Initial Study of Anisotropic Textures for Identification of Blood Vessels in 7T MRI Brain Phase Images

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Phillip Deshawn Barnes, M.S.
Graduate Program in Biomedical Engineering

The Ohio State University
2010

Dissertation Committee:
Dr. Bradley Clymer, Adviser
Dr. Petra Schmalbrock
Dr. Donald Chakeres
Abstract

Within medical science, pattern recognition is the basis for computer-aided diagnosis (CAD), which assists doctors (in particular radiologists) in the interpretation of medical images, whose quality and usefulness are constantly evolving. Modern magnetic resonance imaging (MRI) scans can provide information about subvoxel anatomical structures (e.g. microvessels) that may not be specifically resolvable within a given image set. Hence, subvoxel anatomical structures within a given image may not be readily apparent to the observer (i.e. radiologist); nevertheless, their presence may be detected through their statistical relationship with their surrounding voxels. Such statistical relationships can be characterized by texture features.

The goal of this research dissertation is to investigate the feasibility of using anisotropic texture features for the identification of blood vessels that may not be specifically resolvable in the image datasets. Such features can be used as the basis for the feature extraction component of a complete pattern recognition system for the purpose of automatically identifying blood vessels in the human brain. The approach of this project is to apply 2D statistical texture features as inputs to a classifier (such as a neural network) to analyze MRI images. The specific aims of this dissertation are:
a) to provide a set of texture features extracted from 7T MRI human brain phase images that demonstrate the ability to characterize the presence of underlying microvessel structures; b) to provide a classifier, in particular a neural network architecture that makes use of the extracted features; and c) to evaluate the performance of the feature-classifier combination.

The results of this research demonstrated the feature-classifier combination exhibit reasonably well generalization across the testing data, and suggest it may be possible for a computer to discriminate hidden vessels not detectable by human observers.
Dedication

This dissertation is dedicated to my family and friends. Thank you for everything.
Acknowledgments

This dissertation would not have been possible without the help and guidance of many people. I would like to thank Dr. Bradley Clymer for his indispensable time, help, and guidance. Dr. Clymer put a lot of energy and care into helping me grow as an independent researcher. I cannot thank him enough, but I will try again – thank you. I would like to thank Dr. Petra Schmalbrock and Dr. Donald Chakeres for also serving as members of my PhD committee. Further, I would like to thank Dr. Brent Woods for always being there and taking the time to assist me in honing ideas for experiments.

I would like to thank Amir Abduljalil and Dr. Steffen Sammet for assisting me in obtaining the MRI data sets and blood vessel marked images, respectively, without their help this dissertation would not have been possible. I would also like to thank the many faculty and staff members who positively influence my life, especially Dr. Stephen Lee, Dr. Hooshang Hemami, Dr. Steven Bibyk, Assistant Dean Minnie McGee, Ruby Smith, and Felicia Sawyer.

Through the entire journey I was never alone. My entire family and all of my friends supported me each step of the way, especially my mother, father, sister, brother, and my lovely wife Anishka Turner-Barnes. Thank you all very much!!!
Vita

January 5, 1976 ...............................................Born – Cleveland, Ohio USA

Autumn 1994 ....................................................Began Undergraduate Studies, The Ohio State University

Spring 2001 ....................................................B.S., The Ohio State University

Spring 2004 ....................................................M.S., The Ohio State University

Publications


Fields of Study

Major Field: Biomedical Engineering
Table of Contents

Abstract ........................................................................................................................................ iii

Dedication ....................................................................................................................................... v

Acknowledgments ........................................................................................................................ vi

Vita ................................................................................................................................................ vii

List of Tables .................................................................................................................................. xiii

List of Figures .............................................................................................................................. xv

Chapter 1 ......................................................................................................................................... 1

Introduction .................................................................................................................................... 1

1.1 Pattern Recognition ................................................................................................................. 1

1.2 Motivation .............................................................................................................................. 3

1.3 Problem Statement ................................................................................................................ 5

1.4 Organization .......................................................................................................................... 6

Chapter 2 ....................................................................................................................................... 8

MR Angiography & Susceptibility Imaging ................................................................................... 8

2.1 Magnetostatic Fields .............................................................................................................. 9
<table>
<thead>
<tr>
<th>2.1.1 Electromagnetics Overview</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.2 Magnetic Field Intensity, Susceptibility, &amp; Maxwell’s Equations for Magnetostatics</td>
<td>10</td>
</tr>
<tr>
<td>2.1.3 Behavior of Magnetic Materials</td>
<td>13</td>
</tr>
<tr>
<td>2.1.4 Boundary Conditions for Magnetostatic Fields</td>
<td>14</td>
</tr>
<tr>
<td>2.2 MR Angiography Techniques and Susceptibility Weighted Imaging</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Magnitude and Phase Imaging</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Geometries and Compartmentalization Models</td>
<td>24</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>33</td>
</tr>
</tbody>
</table>

Feature Extraction and Classification

| 3.1 Overview of a Pattern Recognition System                      | 33 |
| 3.2 Preprocessing of Phase Images                                 | 36 |
| 3.3 Feature Extraction: Phase Image Texture Features              | 41 |
| 3.3.1 Phase Image Marginal Statistics                            | 43 |
| 3.3.2 Marginal Statistics of Phase Image Spatial Derivatives      | 45 |
| 3.3.3 Second Order Spatial Partial Derivatives of the Phase Image | 47 |
| 3.3.4 Co-occurrence Texture Measures on Phase Image              | 49 |
3.4 Classification

3.4.1 Feature Selection Process

3.4.2 Assignment of the Neural Network Architecture

Chapter 4

Results and Discussion

4.1 Histogram Analysis

4.1.1 Histogram Analyses of Phase Image Marginal Statistics

4.1.2 Histogram Analyses of Statistics of Phase Image Spatial Derivatives

4.1.3 Histogram Analyses of Co-occurrence Textures on Phase Image

4.2 Correlation Analysis and Complete List of Candidates

4.3 Performance Evaluation of Features and Classifier via ROC Area Index Using Training Data

4.4 Performance Evaluation of Features and Classifier via Color-Coded Images Using Training Data

4.5 Performance Evaluation of Features and Classifier via ROC Area Index and Color-Coded Images Using Testing Data

Chapter 5
Conclusions and Future Work ................................................................. 108

5.1 Summary ............................................................................................... 108

5.2 Conclusions of Research ....................................................................... 110

5.3 Future Work .......................................................................................... 111

Bibliography ................................................................................................ 112
List of Tables

Table 3.1  Descriptors of texture based on intensity histogram, where z is a random variable indicating the phase value, p(z_i) is the histogram of the phase image values in a region, and L is the number of possible gray levels in the image. Note these statistics cannot sense directional differences ............................................................ 44

Table 3.2  1st and 2nd order partial spatial derivatives. .................................................... 48

Table 3.3  The 14 Haralick Features ................................................................................. 52

Table 4.1  Descriptors of texture based on intensity histogram, where z is a random variable indicating the phase value, p(z_i) is the histogram of the phase image values in a region, and L is the number of possible gray levels in the image. Note these statistics cannot sense directional differences ............................................................ 69

Table 4.2  Potential feature candidates of the marginal statistic feature extraction methodology. ................................................................................................. 75

Table 4.3  1st and 2nd order partial spatial derivatives. .................................................... 76

Table 4.4  Potential feature candidates of the H matrix feature extraction methodology 78

Table 4.5  Potential Haralick feature candidates .............................................................. 80

Table 4.6  The encoded numbers and their corresponding Haralick features used in Table 4.5 ............................................................................................................. 81
Table 4.7 List features to be used as candidates for the forward stepwise discriminant feature selection process. ................................................................. 84

Table 4.8 The feature candidates resulting from the forward feature selection process... 85

Table 4.9 Input weight (IW) matrix. Each entry is the value of that connection’s weight from the input node to the corresponding neuron in the hidden layer. ......................... 87

Table 4.10 Hidden layer bias values for corresponding hidden layer neurons. .............. 87

Table 4.11 Output layer weights. Each entry is the value of that connection’s weight from the hidden layer neuron to the output neuron......................................................... 88

Table 4.12 Output layer bias value for the output neuron. ......................................................... 88
List of Figures

Figure 2.1  A typical magnitude image at 7T. ................................................................. 19

Figure 2.2  The corresponding post-processed phase image of Figure 2.1. Note: The post-processing step is explained in Chapter 3 ................................................................. 21

Figure 2.3 Magnitude vs. Phase Image: (a) Typical magnitude image at 7T. Notice the variation of contrast throughout the image, and the difficulty in identifying blood vessels and its surrounding tissue in the two marked regions of low antenna sensitivity. (b) The corresponding post-processed phase image. Notice the corresponding regions and structures in the phase image are easier to distinguish than their magnitude counterpart. Note: The post-processing step is explained in Chapter 3 ............................................. 22

Figure 2.4 Field distributions of $\chi_m$ and $\Delta B$ for a cylinder oriented parallel to the main static field. (a) The field distribution of $\chi_m$. (b) The corresponding field distribution of $\Delta B$ ................................................................. 26

Figure 2.5 Field distributions of $\chi_m$ and $\Delta B$ for a cylinder oriented perpendicular to the main static field. (a) The field distribution of $\chi_m$. (b) The corresponding field distribution of $\Delta B$ ........................................................................................................................................ 27
Figure 2.6 A vessel with an unspecified venous blood volume fraction ($\lambda$) in a voxel. There are two compartments: 1) inside the vessel (int) and 2) outside the vessel but within the voxel (ext).

Figure 2.7 Compartmentalization Cases. (a) The vessel is smaller than a voxel. (b) The vessel is bigger than a voxel.

Figure 3.1 A pattern recognition system can be partitioned into components such as the one shown here. A sensor converts electromagnetic waves, sound, or other physical inputs into signal/image data. The preprocessing step usually involves applying some type of linear or non-linear filtering operation to the data before it moves on to the feature extractor. A feature extractor measures the objects (as represented in the signal/image data) properties that are useful for classification. The classifier uses these features to assign the sensed object into a category.

Figure 3.2 The outline of the steps involved in the preprocessing component of the pattern recognition system: (a) The raw phase image. (b) The result of applying a highpass filter to the raw phase image. (c) The result of applying an autocorrelation filter and then performing histogram equalization.

Figure 3.3 A typical raw phase image and its corresponding highpass filtered version: (a) A raw phase image, notice the large ripples obscuring the brain matter. (b) The resulting highpass filter version of the raw phase image. Notice the result of applying a 3D highpass Gaussian filter allows the brain matter to be more observable.
Figure 3.4 Plots of autocorrelation for various regions of a typical phase image: (a) The plot of the autocorrelation for a region in the air surrounding the head. Notice its resemblance to a Dirac delta function. (b) The plot of the autocorrelation for a region containing only brain tissue. (c) The plot of the autocorrelation for a region containing both blood vessels and surround brain tissue. (d) The plot of the autocorrelation for a region containing both brain tissue and air.

Figure 3.5 The resulting phase image after it has been highpass filtered, ran through the autocorrelation filtering process, been histogram equalized, and finally having its values requantized to integers ranging from 0 to 31.

Figure 3.6 A given input with 8 gray levels (1 through 8) and its corresponding GCLM, defined for the nearest neighbor in the horizontal direction.

Figure 3.7 Four directions of adjacency as defined for calculation of the Haralick texture features: (a) The horizontal direction. (b) The vertical direction. (c) The left diagonal direction. (d) The right diagonal direction.

Figure 3.8 Signal flow graph model of a neuron. $x_0$ through $x_n$ represent the inputs, $w_0$ through $w_n$ represent their associated synaptic weights, $\Phi(\bullet)$ represents the transfer function, and $y$ represents the neuron’s output. Note the input $x_0 = +1$ and the associated synaptic weight $w_o$ is equal to the bias of the neuron.

Figure 3.9 A block diagram representation of a typical neural network.
Figure 3.10 Fully-connected feedforward network with two hidden layers and one output layer

Figure 3.11 Flow Chart outlining the feature selection process.

Figure 3.12 A plot of the histograms of a feature variable $x_1$ and $x_2$: (a) An example of a feature variable ($x_1$) with relatively good discrimination power. For this feature the nonvessel pixels tend to be centered around zero then rapidly decrease in concentration on either side of zero, and extending pass ten. The blood vessel pixels are centered around two, and remains relatively concentrated up to about five and extents all the way pass twenty. (b) An example of a feature variable ($x_2$) with relatively poor discrimination power. Both nonvessel and vessel pixels values are grouped together, and the two distributions almost completely overlap.

Figure 3.13 The general form of the fully-connected feedforward architecture used for this research project.

Figure 4.1 Histogram of the candidate marginal statistics for $W = 3 \times 3$: (a) The mean. (b) The standard deviation.

Figure 4.2 Histogram of the candidate marginal statistics for $W = 5 \times 5$: (a) The mean. (b) The standard deviation (c) The uniformity (d) The entropy.

Figure 4.3 Histograms of candidate marginal statistics for $W = 7 \times 7$: (a) The uniformity. (b) The entropy.
Figure 4.4 Histograms for the candidate marginal statistics for W = 9x9 (a) The uniformity. (b) The entropy. ................................................................. 74

Figure 4.5 The classifier’s architecture consists of 6 inputs, 12 hidden layer neurons, 1 output layer neuron. IW stands for input weights and LW stand for layer weights. ........ 86

Figure 4.6 The ROC curve for the subject 1 uncaffeinated ground truth image data. The ROC area index is 0.7901. ............................................................... 89

Figure 4.7 The ROC curve for subject 1 combined ground truth image data. The ROC area index is 0.7350. ................................................................. 91

Figure 4.8 Uncaffeinated ground truth Vs. Combined ground truth for subject 1.......... 91

Figure 4.9 Color coded image super imposed on anatomical phase image of subject 1 slice 3 using uncaffeinated ground truth...................................................... 94

Figure 4.10 Color coded image super imposed on anatomical phase image corresponding to Figure 4.9 using combined ground truth............................................. 94

Figure 4.11 Zoom-in version of color coded image super imposed on anatomical phase image of subject 1 (a) A zoomed-in version of Figure 4.9. (b) A zoomed-in version of Figure 4.10................................................................. 95

Figure 4.12 The ROC curve for the subject 2 uncaffeinated ground truth image data. The ROC area index is 0.7673 ................................................................. 97

Figure 4.13 The ROC curve for subject 2 combined ground truth image data. The ROC area index is 0.7506. ................................................................. 98
Figure 4.14 Uncaffeinated truth ground Vs. Combined ground truth of subject 2........... 98

Figure 4.15 Color coded image of super imposed on anatomical phase image of subject 2 slice3 using uncaffeinated ground truth................................................................. 99

Figure 4.16 Color coded image super imposed on anatomical phase image corresponding to Figure 4.15 using combined ground truth.......................................................................................................................... 99

Figure 4.17 Zoom-in version of color coded image super imposed on anatomical phase image of subject 2 (a) A zoomed-in version of Figure 4.15. (b) A zoomed-in version of Figure 4.16. ................................................................................................................................. 100

Figure 4.18 The ROC curve for the subject 3 uncaffeinated ground truth image data. The ROC area index is 0.8136. .................................................................................................................................................................. 101

Figure 4.19 The ROC curve for subject 3 combined ground truth image data. The ROC area index is 0.7929. .................................................................................................................................................................. 102

Figure 4.20 Uncaffeinated gound truth Vs. Combined ground truth of subject 3........ 102

Figure 4.21 Color coded image of superimposed on anatomical phase image of subject 3 slice1 using uncaffeinated ground truth............................................................... 103

Figure 4.22 Color coded image super imposed on anatomical phase image corresponding to Figure 4.27 using combined ground truth.......................................................... 103

Figure 4.23 Zoom-in version of color coded image super imposed on anatomical phase image of subject 3 (a) A zoomed-in version of Figure 4.21. (b) A zoomed-in version of Figure 4.22 ........................................................................................................................................................................ 104
Figure 4.24 Comparison of subjects 1, 2 and 3 using uncaffeinated ground truth. ....... 105
Figure 4.25 Comparison of subjects 1, 2 and 3 using combined ground truth. ............... 105
1.1 Pattern Recognition

Pattern recognition can be defined as the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns (Jain, Dun and Mao 2000). A pattern recognition system generally consists of three major components: 1) a sensor that gathers the observations to be classified, 2) a feature extraction mechanism that computes the numeric or symbolic information from the observations, and 3) a classifier scheme that does the actual job of classifying the observations, relying on the extracted features (Duda, Hart and Stork 2001). Pattern recognition is used in diverse applications: handwriting algorithms, financial analysis, gene expressions, biometrics, and in medical imaging. Within medical science, pattern recognition is the basis for computer-aided diagnosis (CAD), which assists doctors (in particular radiologists) in the interpretation of medical images. Medical imaging techniques such as x-ray, magnetic
resonance imaging (MRI), ultrasound, etc., can provide a great deal of anatomical and physiological information about clinical patients.

The quality and usefulness of the medical imaging techniques is constantly evolving, leading to better patient care. For example, modern imaging techniques based on MRI can provide 2D, 3D, and even 4D datasets consisting of millions, billions, even trillions of pixels and/or voxels. Such scans can also provide information about subvoxel anatomical structures (e.g. microvessels) that may not be specifically resolvable within a given image set. Hence, subvoxel anatomical structures within a given image may not be readily apparent to the observer (i.e. radiologist); nevertheless, their presence may be detected through their statistical relationship with their surrounding voxels. Such statistical relationships can be characterized by texture features. For purposes of the current discussion, texture is defined to mean an attribute representing the statistical spatial arrangement of gray levels of the pixels in a region of an image (Castleman 1996). These texture features and the methods used to compute them form the basis of the feature extraction mechanism of the pattern recognition system that is investigated during the course of this dissertation research. In particular, for this dissertation the sensor is a 7T MRI machine, the classifier is a neural network, the patterns of interest are veins embedded in surrounding brain tissue, and the features are texture measures extracted from these patterns.
1.2 Motivation

Noninvasive assessment of vasculature is of great importance, because vasculature is implicated in many common medical problems including tumors, stroke, and hypertension, diabetes mellitus, vasculitis, and venous thrombosis. Vasculature of the brain is organized in a complex fashion, ranging in diameters from a few microns to several millimeters with flow velocities ranging from mm/s in the capillaries to several cm/s in arteries. From a clinical point of view, it is necessary to differentiate between arterial and venous vasculature, as well as to have the capabilities for depicting vessels of a wide range of sizes. While several excellent standard magnetic resonance (MR) angiography techniques (e.g. time-of-flight and phase contrast) are available for depicting arteries larger than a few tenths of millimeters, most of these techniques are less sensitive for depicting veins (Haacke, et al. 1999). Furthermore, although capillary flow may be assessed indirectly through MR perfusion techniques, there is no clinically accepted standard for visualizing small vessels ranging from 50-200 microns within the human brain.

Reichenbach (Reichenbach, Venkatesan, et al. 1997) have shown that small veins can be depicted in T2* weighted, long TE gradient echo imaging at 1.5T and 3T. Veins appear dark in these images based on the paramagnetic properties of deoxyhemoglobin. This so-called blood oxygen level dependent (BOLD) effect (Ogawa, Lee and Nayak 1990), can be further emphasized in phase rather than magnitude reconstructed images (Reichenbach and Haacke 2001), whose physical principles are discussed in more detail in Chapter 2. Based on the estimated magnetic susceptibility for venous blood of
specified oxygen fraction and hemocrit and TE of 40ms at 1.5T, it was estimated that veins of about 200 micron diameter can be seen with 0.5x0.5mm in-plane resolution, i.e. the vessel occupies only about 1/10 of the image voxel volume (Reichenbach and Haacke 2001). While spatial resolution of 0.23x0.23x2.0mm voxels were achieved at 1.5T allowing for depiction of even smaller vessels, the long TE needed at 1.5T and the low signal-to-noise ratio (SNR) required an excessive scan time of 1.5 hours (Reichenbach, et al. 1998).

Since susceptibility effects increase linearly with magnetic field strength, the BOLD effect is much more pronounced at 7T. This allows shorter TE (10-15ms compared to 30-40ms at 1.5T) and thus shorter scan times, while at the same time providing increased SNR. Despite these exciting new developments, extensive work needs to be completed to improve the overall image quality at ultra high field. Presently there are serious image quality issues related to ultra high field MRI. These include inhomogeneity of antenna sensitivity, resulting in variable SNR and image contrast within the same image, and severe magnetic susceptibility artifacts (Ibrahim, et al. 2001). Although the variable image intensity and SNR does not present an insurmountable problem for a radiologist identifying microvessels (of sufficient diameter), they lead to marked limitations in the applicability of standard image processing tools for feature extraction and classification for ultra high field MRI identification of microvasculature.

Although these artifacts are not an insurmountable problem for radiologists identifying microvessels, in the presence of such image quality issues manual identification is prone to both inter- and intra-observer variations. For example,
controlling a mouse with the precision to highlight identifiable microvessels on a computer monitor is extremely difficult for even the most trained radiologist. Also as mentioned above, some subvoxel microvessels are not readily identifiable within images, and thus will remain unidentified. Clinicians and researchers would benefit from a CAD pattern recognition system that provides reliable automatic identification of microvessels.

1.3 Problem Statement

There is a susceptibility difference between fully oxygenated and fully deoxygenated blood red blood cells due to the presence of deoxyhemoglobin’s unpaired electrons, and it is this susceptibility difference that is exploited in susceptibility weighted imaging (SWI) to differentiate veins from arteries and the surrounding tissue. SWI is an active and ongoing area of research at institutions including The Ohio State University. The goal of this research dissertation is to investigate the feasibility of using anisotropic texture features for the identification of blood vessels that may not be specifically resolvable in the image datasets. Such features can be used as the basis for the feature extraction component of a complete pattern recognition system for the purpose of automatically identifying blood vessels in the human brain. The approach of the project is to apply 2D statistical texture features as inputs into a classifier (such as a neural network) to analyze MRI images. The specific aims of this dissertation are: a) to provide a set of texture features extracted from 7T MRI human brain phase images that demonstrate the ability to characterize the presence of underlying microvessel structures; b) to provide a classifier,
in particular a neural network architecture that makes use of the extracted features; and c) to evaluate the performance of the feature-classifier combination.

The performance of the feature-classifier combination for this project is demonstrated with human brain 7T MRI phase images. Classification is performed on a pixel-by-pixel basis. In this study there are two categories: “blood vessel” or “not blood vessel” (e.g. surrounding brain tissue). The evaluation of the classifier is performed using manual highlighted images provided a radiologist. The performance is measured using two methods. The first method involves computing the receiver operating characteristic (ROC) curve for the classifier. The second method is more qualitative; it involves creating color-coded output images that provide a visual method of determining the feature-classifier combination’s performance. We superimpose color-coded images on the anatomical 7T phase images. The objective of this research is to perform a preliminary study to show feasibility, and therefore we do not claim a completed pattern recognition system or toolkit to be used by radiologist, but begin the necessary research into statistical features for the purpose of identifying microvessel structures in the brain.

1.4 Organization

In Chapter 2 of this dissertation, the basic physical principles behind MR angiography and susceptibility weighted imaging are discussed. Chapter 3 provides an overview of the pattern recognition system being investigated in this dissertation, and discusses the mathematical principles governing the anisotropic texture features. In Chapter 4 we present the evaluation results of the classifier’s performance. Chapter 5
offers the conclusions of this research, discusses the significance of the results, and provides an overview of the methods that can be employed to extend this research.
This chapter is an overview of the basic principles behind magnetic resonance (MR) angiography and susceptibility weighted imaging. In particular this chapter is divided into four subsections. Section 2.1 provides a working definition of magnetic susceptibility and its ability to characterize magnetic material. Section 2.2 reviews traditional MR angiographic methods and their inadequacies for the purposes of this research project, and begins a discussion of the bulk magnetic susceptibility (BMS) effect. Section 2.3 reviews the basic principles involved in MR imaging, and the advantages of using phase imaging as opposed to magnitude imaging. Finally Section 2.4 presents the relationship between local variations in magnetic susceptibility and the local variations in the static magnetic field for cylindrical geometries, and an overview of the partial volume effects.
2.1 Magnetostatic Fields

2.1.1 Electromagnetics Overview

Stated simply, electromagnetics is the study of the effects of electric charges at rest and in motion. There are two kinds of charges: positive and negative. Both positive and negative charges are sources of an electric field. Moving charges produce a current, which gives rise to a magnetic field. A field is a spatial distribution of a scalar or vector quantity, which may or may not be a function of time. A time varying electric field is accompanied by a magnetic field, and vice versa. Thus time varying electric and magnetic fields are coupled, resulting in an electromagnetic field.

There are four fundamental vector field quantities in electromagnetics: electric field intensity \((E)\), electric flux density (or electric displacement) \((D)\), magnetic flux density \((B)\), and magnetic field intensity \((H)\). Material properties determine the relationships between \(E\) and \(D\) and between \(B\) and \(H\) (Cheng 1989). For cases involving time-dependent field variations the electric and magnetic field quantities are coupled; that is, time-varying \(E\) and \(D\) will give rise to \(B\) and \(H\), and vice versa. However, when there is no time variation (i.e. the static case), the electric field quantities \(E\) and \(D\) are uncoupled from quantities \(B\) and \(H\).

In the electromagnetic model there are three universal constants, in addition to the above four field quantities, and they relate to the properties of free space (a vacuum). The universal constants are as follows: speed of electromagnetic waves in free space \((c)\); permittivity of free space \((\varepsilon_0)\); and permeability of free space \((\mu_0)\). The constants \(\varepsilon_0\) and \(\mu_0\) pertain to electric and magnetic phenomena in a vacuum, respectively: \(\varepsilon_0\) is the...
proportionality constant between the electric flux density $D$ and the electric field intensity $E$ in free space, as stated in Equation 2.1

\[ D = \varepsilon_0 E \]  

(Equation 2.1)

and $\mu_0$ is the proportionality constant between the magnetic flux $B$ and the magnetic field intensity $H$ in free space, as stated in Equation 2.2.

\[ B = \mu_0 H \]  

(Equation 2.2)

In all of the discussion that follows we are only interested in static cases, in particular magnetostatics (steady magnetic fields). The magnetic flux density $B$ is the only vector quantity needed in discussing magnetostatics in free space and is related to the magnetic force acting on a charge moving with a given velocity. The magnetic field intensity vector $H$ is useful in the study of magnetic fields in material media.

2.1.2 Magnetic Field Intensity, Susceptibility, & Maxwell’s Equations for Magnetostatics

According to the elementary atomic model of matter, all materials are composed of atoms, each with a positively charged nucleus due to the presence of proton(s) and a number of orbiting negatively charged electrons. The orbiting electrons cause circulating currents and form microscopic magnetic dipoles. In addition, both the electrons and the nucleus of an atom rotate (spin) on their own axes with certain magnetic dipole moments.
In the absence of an external magnetic field the magnetic dipoles of the atoms of most materials (except permanent magnets) have random orientations, resulting in no net magnetic moment. The application of an external magnetic field causes both the alignment of the magnetic moments of the spins and an induced magnetic moment due to a change in the orbital motion of electrons, thereby magnetizing the material. The strength of this magnetizing effect is measured by the magnetization vector \( \mathbf{M} \) (Cheng 1989). The macroscopic effect of magnetization can be studied by incorporating the equivalent volume current density \( \mathbf{J}_m \), expressed in Equation 2.3, Maxwell’s equation corresponding to Ampere’s law (see Equation 2.4), that gives us a definition of the \( \mathbf{H} \) field (Equation 2.5).

\[
\mathbf{J}_m = \nabla \times \mathbf{M} \quad \text{(Equation 2.3)}
\]
\[
\nabla \times \mathbf{B} = \mu_0 \mathbf{J}_m \quad \text{(Equation 2.4)}
\]
\[
\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M} \quad \text{(Equation 2.5)}
\]

When the magnetic properties of the material are linear, the magnetic field and the magnetization are proportional to each other and can be described by Equation 2.6

\[
\mathbf{M} = \chi_m \mathbf{H} \quad \text{(Equation 2.6)}
\]
where $\chi_m$ is a dimensionless quantity called magnetic susceptibility. Substitution of Equation 2.6 into Equation 2.5 yields the relation between $B$ and $H$, as well as the relation between $\mu$ and $\chi_m$.

$$B = \mu_o(1 + \chi_m)H = \mu_o\mu_r H = \mu H \quad \text{(Equation 2.7a)}$$

$$B = ((1+\chi_m)/\chi_m)\mu_o M \quad \text{(Equation 2.7b)}$$

or

$$H = (1/\mu)B \quad \text{(Equation 2.8)}$$

where $\mu_r$ is another dimensionless quantity known as the relative permeability of the material, and is related to $\chi_m$ and $\mu_o$ as expressed in Equations 2.9 and 2.10.

$$\mu_r = 1+\chi_m = \mu/\mu_o \quad \text{(Equation 2.9)}$$

or

$$\chi_m = \mu/\mu_o - 1 \quad \text{(Equation 2.10)}$$

The parameter $\mu = \mu_o\mu_r$ is the absolute permeability (or sometimes just permeability) of the material. The magnetic susceptibility is the degree of magnetization of a material in response to an applied magnetic field, and can be a function of space coordinates in media with a spatial distribution of magnetic dipoles. The two fundamental governing differential Maxwell’s equations for magnetostatics are stated in Equations 2.11 and 2.12, whose significance to the present work is discussed shortly.
\[ \nabla \cdot \mathbf{B} = 0 \quad \text{(Equation 2.11)} \]
\[ \nabla \times \mathbf{H} = \mathbf{J} \quad \text{(Equation 2.12)} \]

2.1.3 Behavior of Magnetic Materials

The following explanation is a summary of material presented in (Cheng 1989).

In Equation 2.6 we have described the macroscopic magnetic property of linear material by defining susceptibility \( \chi_m \). For the purpose of this discussion (and work involved in this dissertation), we limit the discussion and classification of magnetic materials into two groups: diamagnetic and paramagnetic. The other types/classifications of magnetic materials are omitted because they are usually avoided in clinical MRI. A material is said to be

- Diamagnetic, if \( \chi_m \) is a very small negative number (Typically measured in parts per million)

- Paramagnetic, if \( \chi_m \) is a very small positive number (Typically measured in parts per million).

In diamagnetic materials, the net magnetic moment due to the orbital and spinning electrons in any particular atom is zero in the absence of an externally applied magnetic field. The application of an external magnetic field produces a force on the orbiting electrons, causing a perturbation in the angular velocities, resulting in a net magnetic moment being created. The induced magnetic moments within diamagnetic materials
always oppose the applied field, thus reducing the \( B \) field. Diamagnetic materials exhibit no permanent magnetism, and the induced magnetic moment disappears when the applied field is withdrawn.

In paramagnetic materials, the magnetic moments due to the orbiting and spinning electrons do not cancel completely, and atoms and molecules have a net average magnetic moment. An externally applied magnetic field, in addition to causing a very weak diamagnetic effect, tends to align the molecular magnetic moments in the direction of the applied field, thus increasing the \( B \) field. The alignment process is, however, impeded by random thermal vibrations. There is little coherent interaction, and the increase in the \( B \) field is quite small. Paramagnetism arises mainly from the magnetic dipole moments of the unmatched spinning electrons.

2.1.4 Boundary Conditions for Magnetostatic Fields

In order to understand or solve problems concerning magnetic fields in regions having media with different magnetic properties, it is necessary to study/understand the boundary conditions that \( B \) and \( H \) vectors must satisfy at the interfaces. The magnetostatic boundary conditions are established by applying Equations 2.11 and 2.12 to a small pillbox and a small closed path, respectively, which include the interface (Cheng 1989). From the divergenceless nature of the \( B \) field in Equation 2.11 we may conclude directly that the normal component of \( B \) is continuous across an interface, that is,
Normal to a surface: \[ B_{1n} = B_{2n} \] (Equation 2.13)

The boundary condition for the tangential components of the magnetostatic field is obtained from the integral form of the curl equation for \( \mathbf{H} \), Equation 2.12, which results in,

\[
H_{1t} - H_{2t} = J_{sn}
\] (Equation 2.14)

where \( J_{sn} \) is the surface current density on the interface normal to the enclosing contour. However, when the conductivities of both media (materials) are finite, currents are defined by volume current densities and free surface currents do not exist on the interface. Hence, \( J_{sn} \) equals zero, and the tangential component of \( \mathbf{H} \) is continuous across the boundary of almost all physical media; it is discontinuous (with the amount of discontinuity being determined in Equation 2.14) only when an interface with an ideal perfect conductor or a superconductor is assumed. Thus for our purposes \( J_{sn} \) is considered to be zero, resulting in the second boundary condition (Equation 2.15).

Tangential to Surface: \[ H_{1t} = H_{2t} \] (Equation 2.15)

As mentioned above, the susceptibility, and therefore the static \( \mathbf{B} \) field can be a function of space coordinates (and for our experiments they are). As predicted by the fundamental governing equations (Equations 2.11 and 2.12) including the fact that the boundary
conditions (Equations 2.14 and 2.15) must always be satisfied as we move region to region, a step change in $\chi_m(x,y,z)$ can cause perturbations in $B(x,y,z)$ that are larger than the size of the object, depending on its shape and orientation. This point is further explained over the course of the next three sections when we discuss magnetic resonance angiography, phasing imaging, and compartmentalization.

2.2 MR Angiography Techniques and Susceptibility Weighted Imaging

MR angiography is an important diagnostic tool for the depiction and characterization of blood vessels and blood flow in the human brain. In regard to clinical field strengths (e.g. 1.5T and 3T), to date the most widely used techniques to create angiograms are time-of-flight (TOF) and phase-contrast MR angiography. Both techniques require the presence of moving blood: The former is sensitive to rapidly inflowing spins whereas the latter represents a proton velocity map. TOF methods use the inflow of fresh, unsaturated blood so that the vessels appear bright compared with the surrounding brain tissue. TOF sequences are usually flow compensated in the section-select and read out directions to avoid spin dephasing (Reichenbach, Venkatesan, et al. 1997). Phase-contrast angiographic methods make use of at least two images, one of which is velocity compensated, whereas the other is velocity sensitized in one direction. Subtraction is then used to suppress background tissue and extract vessel structures and velocity of blood within the vessels (Reichenbach, Venkatesan, et al. 1997). However, SNR and contrast limit the above techniques to relatively fast flow in larger vessels; therefore these methods have limitations in terms of the ability to visualize slow flowing
blood in small vessels. To overcome the problems associated with slow flow, one can exploit the magnetic properties of blood. There is a susceptibility difference between fully oxygenated and fully deoxygenated blood red blood cells due to the presence of deoxyhemoglobin’s unpaired electrons, and it is this susceptibility difference that is exploited in what is called susceptibility weighted imaging (SWI) to differentiate veins from arteries and the surrounding tissue.

The bulk volume susceptibility shift effect of deoxygenated blood causes a frequency shift in the spins of the protons, which leads ultimately to a phase difference between the signal detected from venous blood spins and the surrounding tissue spins. Thus, the underlying principle that makes susceptibility imaging possible is the frequency shift caused by the bulk magnetic susceptibility (BMS) effect that arises from compartmentalization of the paramagnetic species, such as the deoxyhemoglobin molecules in a vein. Therefore, susceptibility differences between a vein and its surrounding tissue cause local magnetic field differences, which lead to differences in local precession frequencies that manifest as phase differences for spins inside a venous vessel with respect to the surrounding tissue. These phase differences can be used as contrast in phase images. Therefore local deoxygenation content can be used as an endogenous contrast agent.

It is important to re-emphasize that there is a relationship between the local magnetic field changes ($\Delta B$) and the local variations in susceptibility ($\Delta \chi$), and such a relationship enables the use of phase information to provide insight into magnetic properties of tissues relative to each other (as explained in the next section). Obtaining
an expression of the relation between $\Delta B(x,y,z)$ and $\Delta \chi(x,y,z)$ involves solving partial
differential equations with boundary conditions, and for certain geometries there exist
closed form solutions. Such closed form solutions are provided in Section 2.4. In the
next section we discuss in more detail the relationship between $\Delta B(x,y,z)$ and $\Delta \chi(x,y,z)$, as
well as imaging methodologies used to exploit the BMS effect.

2.3 Magnitude and Phase Imaging

Magnetic resonance imaging (MRI) is an imaging technique used primarily in
medical settings to produce high quality images of the inside of the human body. MRI is
based on the interaction of a nuclear spin with an external magnetic field $B_o$. The
dominant nucleus in MRI is the proton in hydrogen, and its interaction with the external
magnetic field results in the precession of the proton spin about the $B_o$ field direction,
leading to what are termed magnetic dipoles. The precession angular frequency for the
proton magnetic moment vector (and for the spin axis as well) is given by Equation 2.16
(Haacke, et al. 1999)

$$\omega_o = \gamma B_o$$

(Equation 2.16)

where $\gamma$ is a constant called the gyromagnetic ratio. This precession frequency is referred
to as the Larmor frequency, and Equation 2.16 is referred to as the Larmor equation.

The MR signal is the result of excitation of the individual magnetized protons
(spinning dipoles) within the object by irradiation with radiofrequency (RF) energy of a
specific frequency. This energy absorption causes the displacement of magnetic moments from equilibrium, resulting in an excited system. As the system returns to equilibrium, MR signals are emitted in proportion to the number of excited protons in the sample. The signal is also related to the localized magnetic environment that the protons experience. The detection and acquisition of the signals constitute the basic information necessary for MR imaging. Of particular interest is the fact that the detected MR signal is a complex signal possessing both magnitude and phase components. Spatial signal magnitude variation is caused by both the dipole spatial distribution within the imaged object, and by the spatial sensitivity of antennae used for both the excitation of spinning dipoles and the detection of the emitted RF energy waves. Figure 2.1 illustrates a typical magnitude image acquired using a 7T MRI machine.

Figure 2.1 A typical magnitude image at 7T.
Magnetic susceptibility is an intrinsic tissue property, and local variations in magnetic susceptibility $\Delta \chi(x,y,z)$ cause local variations in the static magnetic field $\Delta B_o(x,y,z)$. Local variations in the static field, $\Delta B_o(x,y,z)$, can be observed by measuring the corresponding local variations in the precession angular frequency $\omega_o$ (refer to Lamor Equation) of the respective spins. Thus $\Delta B_o(x,y,z)$ due to $\Delta \chi(x,y,z)$ causes frequency modulation of RF waves based on spin locations in the $\Delta B_o(x,y,z)$ map. Frequency differences lead to phase differences $\Delta \phi(x,y,z)$ at detection times as expressed by Equation 2.17 (Haacke, et al. 1999).

$$\Delta \phi(x,y,z) = - \gamma * \Delta B(x,y,z) * \text{TE}$$

(Equation 2.17)

where TE is the time delay between spin excitation and signal sampling. Therefore the phase image of an object is a measure of the $\Delta B_o(x,y,z)$ map of that object and indirectly is an indication of the spatial differences in $\chi_m$ in the body. Figure 2.2 illustrate the corresponding post-processed phase image of Figure 2.1.
Figure 2.2  The corresponding post-processed phase image of Figure 2.1.  Note: The post-processing step is explained in Chapter 3.

Phase images are more robust than magnitude images when it comes to dealing with variable signal-to-noise ratio (SNR) that accompany ultra high field MRI scans. The variable SNR results from the difficulty involved in establishing a uniform antenna sensitivity over/around the entirety of the human head, and phase images are less dependent on signal strength than magnitude images, as shown in Figures 2.2 and 2.3. This is analogous to the advantage that frequency modulation has over amplitude modulation used in broadcast radio.
Figure 2.3 Magnitude vs. Phase Image: (a) Typical magnitude image at 7T. Notice the variation of contrast throughout the image, and the difficulty in identifying blood vessels and its surrounding tissue in the two marked regions of low antenna sensitivity. (b) The corresponding post-processed phase image. Notice the corresponding regions and structures in the phase image are easier to distinguish than their magnitude counterpart. Note: The post-processing step is explained in Chapter 3.
For magnitude imaging, in addition to antenna sensitivity, contrast is governed mainly by the intrinsic tissue parameters such as proton density, and inherent T1, T2, or T2* relaxation times. These relaxation times are time constants that represent the tissue properties that relate to the emitted RF signal decay (Haacke, et al. 1999). Whereas (as stated above) the contrast in phase images depends on differences in local precession frequencies, which in turn depend on BMS of different tissue types or different levels of blood oxygenation. This means that with a single MR measurement one in fact obtains two images that bear non-redundant information. Although phase information can be obtained with every imaging sequence, it often remains underused (or even unused). Reasons for the underuse of phase images, namely banding artifact issues (which have been removed from the phase images in Figures 2.2 and 2.3), and methods for overcoming such issues are discussed at length in Chapter 3.

Since susceptibility effects increase linearly with the magnetic field strength as shown in Equation 2.6, the BMS effect is much more pronounced at 7T than at 1.5T. This allows far shorter TE and thus shorter scan times, while at the same time providing significantly increased SNR. Consequently, 7T images far exceed the resolution capabilities of 1.5T images. Despite the advantages of generating phase images using a 7T machine, extensive work still needs to be completed to improve the inherent image quality issues. The issues include inhomogeneous antenna sensitivity, resulting in variable SNR and magnitude image contrast on the same image, and severe magnetic susceptibility artifacts (Ibrahim, et al. 2001). Although the variable image intensity and SNR do not present an insurmountable problem for a radiologist identifying
microvessels, they do lead to marked limitations in applicability of standard image processing tools for automatic identification and subsequent quantification of ultra high field MRI microvasculature.

This research project proposes the use of image analysis techniques that exploit the inherent anisotropic structure of blood vessels for overcoming the previously mentioned imaging issues. The proposed anisotropic imaging techniques are discussed in the next chapter. In the next section we complete our discussion of the relationship between $\Delta B(x,y,z)$ and $\Delta \chi(x,y,z)$ by presenting formulas that model their behavior, and discuss other issues (e.g. mixture effects) that contribute to the previously stated image processing limitations.

2.4 Geometries and Compartmentalization Models

Models for susceptibility variations due to simple tissue compartment geometries are the topic of this section. Such models lead to a better understanding of Equation 2.17 and the relationship between $\Delta B(x,y,z)$ and $\Delta \chi(x,y,z)$. Since we are ultimately interest in identifying blood vessels and blood vessels can be effectively modeled as cylinders, this section focuses exclusively on cylindrical geometries and their orientation within the main magnetic field ($B_0$). When blood vessels are placed within $B_0$, the magnetic field responds differentially inside the lumen of the blood vessels and the surrounding tissue. The difference field ($\Delta B$) is due to the BMS effects discussed in Section 2.2. For the sake of discussion let’s focus on two orientations of a given blood vessel within $B_0$, namely, a parallel and perpendicular orientation. For each case we are modeling the
blood vessel as a cylinder, thus the term blood vessel and cylinder may be used interchangeably. We also define difference field as $\Delta B = B_q - B_o$, with $B_q$ denoting the compartmental fields ($q = \text{int}$ means internal, and $q = \text{ext}$ means external).

For the case of a parallel orientation defined as the blood vessel’s longitudinal axis being parallel to the $B_o$ vector, $B_{\text{ext}} = \mu_{\text{ext}} H_o = \mu_o (1 + \chi_{m,\text{ext}}) H_o$. Thus $H_o$ is only tangent to the surface of the blood vessel, and hence there is no normal component of $B_o$ present. Therefore there is no distortion of $B_o$ outside (external) to the vessel, because the step change in $\Delta B$ aligns with the step change in $\chi_m$ as shown in Figure 2.4.

For the case of a perpendicular orientation defined as the blood vessel’s longitudinal axis being perpendicular to the $B_o$ vector, $H_o$ can have a normal component to the blood vessel surface and hence $B_o$ can also have a normal component to the surface. Also the tangential component of $H_o$ varies for this case around the cross section of the blood vessel due to geometry. Furthermore, the normal components of $B_o$ and $H_o$ depend on their locations on the circular boundary. Thus $\nabla \cdot B = 0$ and $\nabla \times H = 0$ causes a distortion of $B_o$ outside the vessel as shown in Figure 2.5.
Figure 2.4 Field distributions of $\chi_m$ and $\Delta B$ for a cylinder oriented parallel to the main static field. (a) The field distribution of $\chi_m$. (b) The corresponding field distribution of $\Delta B$. 
Figure 2.5 Field distributions of $\chi_m$ and $\Delta B$ for a cylinder oriented perpendicular to the main static field. (a) The field distribution of $\chi_m$. (b) The corresponding field distribution of $\Delta B$.

Therefore the distortion of $\Delta B$ depends on the orientation of the vessel relative to the direction $B_0$. The broadening of $\Delta B$ in Figure 2.5 due to the complicated normal and tangential boundary conditions of $B$ and $H$, and can be larger than the size of the vessel, allowing the presence of small vessels to be observed. For cylinders, the relationship
between $\Delta B(x,y,z)$ and $\Delta \chi(x,y,z)$ has a closed form solution. Using the case of venous blood, the susceptibility difference between a vein and the surrounding tissue can be expressed as

$$\Delta \chi = 4\pi \cdot \Delta \chi_{do} \cdot Hct \cdot (1 - Y) \quad \text{(Equation 2.18)}$$

where, $\Delta \chi_{do} = 0.18 \times 10^{-6}$ (in CGS units) is the susceptibility difference between deoxygenated and fully oxygenated red blood cells, $Y$ is the fraction of oxygen saturation of the blood in the vessel ($0 < Y < 1$), and $Hct$ is the average volume fraction of hematocrit in blood ($Hct \sim 0.41$ to $0.46$ for healthy subjects) (Reichenbach and Haacke 2001). In Figures 2.4 and 2.5 the various intravascular and extravascular fields’ geometric dependence causes corresponding frequency shifts. As governed by the magnetostatic equations and boundary conditions (see Equations 2.11, 2.12, 2.13, and 2.15) the fields for the two compartments can be expressed analytically as (Reichenbach and Haacke 2001)

$$B_{\text{int}} = \left[ 1 + \frac{\chi_{\text{ext}}}{3} + \frac{\chi_{\text{int}} - \chi_{\text{ext}}}{6} \cdot (3\cos^2 \theta - 1) \right] \cdot B_0 \quad \text{(Equation 2.19)}$$

$$B_{\text{ext}} = \left( 1 + \frac{\chi_{\text{ext}}}{3} \right) \cdot B_0 + \left( \frac{\chi_{\text{int}} - \chi_{\text{ext}}}{2} \cdot \sin^2 \theta \cdot \frac{R^2}{r^2} \cdot \cos 2\phi \right) \cdot B_0 \quad \text{(Equation 2.20)}$$
where $\Theta$ is the angle between the vessel axis and the direction of $B_0$, $R$ is the radius of the vessel, $r$ is the perpendicular distance of the (external) spin from the vessel axis, and $\phi$ is the angle between $r$ the plane defined by $B_0$ and the vessel axis. Local magnetic field changes ($\Delta B$) pertaining to the vessel lumen relative to tissue outside the vessel lumen ($r >> R$) can be given as

$$\Delta B = \frac{\Delta \chi}{2} \cdot \left( \cos^2 \theta - \frac{1}{3} \right) \cdot B_0 \quad \text{(Equation 2.21)}$$

The phase difference for spins inside a venous vessel with respect to the surrounding tissue develops according to Equation 2.17. In equation 2.17, the strength of the $\Delta B$ effect increases linearly with the main/static magnetic field strength $B_0$ as shown in Equation 2.21. Thus increasing the applied static magnetic field strength (e.g. from 1.5T to 7T) increases $\Delta B$, which leads to greater phase differences/contrast.

Based on two-compartment models containing blood and tissue (see Figure 2.6), partial volume effects can cause signal cancellation if the protons in the venous blood and the tissue are out of phase for an appropriately chosen TE, thus resulting in (small) veins appearing dark in phase images. Due to the paramagnetic property of deoxyhemoglobin in the veins, there is a phase difference between spin inside the veins and those within the surround brain tissue. This phase difference in conjunction with increased SNR provided by higher field strength magnets allows subvoxel vascular structures to be visualized. In particular, for a given voxel containing small veins plus its surrounding tissue has a different magnetic susceptibility compared to its neighboring voxels containing only
surrounding tissue. These small susceptibility differences can cause enough contrast to allow the presence of the subvoxel vessels to be detected without the use of exogenous contrast agents. This is aided by the $\Delta B$ broadening effect illustrated in Figure 2.5.

Figure 2.6 A vessel with an unspecified venous blood volume fraction ($\lambda$) in a voxel. There are two compartments: 1) inside the vessel (int) and 2) outside the vessel but within the voxel (ext).

Continuing with the compartmentalization discussion there are two cases of particular interest. The first case is when the vessel’s diameter is smaller than the voxels illustrated in Figure 2.7a, and the second case is when the vessel’s diameter is bigger than the voxels illustrated in Figure 2.7b. Case 2 is easier for identifying blood vessels and applying automatic image segmentation algorithms, because the voxels residing completely within the vessel have phase values that are considerably
different from the phase values of the voxels residing completely outside the vessel. For case 2 the vessels can be identified/segmented using simple image thresholding techniques.

Figure 2.7 Compartmentalization Cases. (a) The vessel is smaller than a voxel. (b) The vessel is bigger than a voxel.
For case 1, there is a mixture effect and image segmentation is more difficult. Some voxels contain the phase contributions from both the vessel lumen spins and the surrounding tissue spins. The relative contributions and thus the voxel’s resulting phase value depend on the vessel volume fraction, $\lambda$, (i.e. the amount of the voxel’s volume occupied by vessel) and the vessel orientation to the static field $B_0$. Case 1 (i.e. subvoxel vessels) does not lend itself readily to the use of simple image thresholding techniques; however, such cases are very important to us. In the next chapter we discuss image analysis techniques that attempt to allow us to automatically segment regions within phase images containing both case 1 and case 2.
Chapter 3

Feature Extraction and Classification

3.1 Overview of a Pattern Recognition System

Within medical science, pattern recognition is the basis for computer-aided diagnosis (CAD) systems. CAD describes a procedure that supports the doctor’s interpretations and findings. Pattern recognition can be described as the act of taking in raw data and taking an action based on the category of the pattern (Duda, Hart and Stork 2001). A complete pattern recognition system (refer to Figure 3.1) consists of a sensor that gathers the observations to be classified or described, a feature extraction mechanism that computes numeric or symbolic information from the observations, and a classifier or description scheme that does the actual job of classifying or describing observations, relying on the extracted features. To understand the problem of designing such a system, it is helpful to understand the responsibilities and problems that each component has.
The input into a pattern recognition system is usually a physical (analog) signal (e.g. electromagnetic waves, sound, or temperature) and the sensor (also called a transducer) converts the input signal into digital signal/image data. For this research project, the inputs are RF energy waves emitted from spinning magnetic dipoles (hydrogen protons) within the human brain, and the sensor is a 7T MRI machine whose physical principles have been discussed in Chapter 2. The RF waves are converted by the 7T MRI machine into complex image data, which can be converted into both magnitude images and phase images more pertinent for this research project. The phase images discussed in Chapter 2 are preprocessed by the application of a Gaussian high pass filter and histogram equalization. These preprocessing steps help make the phase images more amenable to feature extraction. For this research project the predominant features of interest are textures, in particular anisotropic texture features, because the patterns we wish to classify (i.e. blood vessels embedded within human brain tissue) are inherently anisotropic structures. The task of the classifier component is to use the feature vector provided by the feature extractor to assign the object to a category (i.e. decide whether a given pixel belongs to a blood vessel or to its surrounding brain tissue). The physical principles underlying the input and sensor components have been discussed in Chapter 2. The remainder of this chapter delves deeper into each of the other above stated pattern recognition components as they relate to this research project.
Figure 3.1 A pattern recognition system can be partitioned into components such as the one shown here. A sensor converts electromagnetic waves, sound, or other physical inputs into signal/image data. The preprocessing step usually involves applying some type of linear or non-linear filtering operation to the data before it moves on to the feature extractor. A feature extractor measures the objects (as represented in the signal/image data) properties that are useful for classification. The classifier uses these features to assign the sensed object into a category.
3.2 Preprocessing of Phase Images

The steps involved in the preprocessing components of the pattern recognition system used in this research project are outlined in Figure 3.2, and this section is dedicated to discussing each of the steps illustrated in the figure. Unprocessed phase images are overwhelmed by large ripples resulting from imperfect shim and copious susceptibility differences and corresponding magnetic field variations at large tissue boundaries (e.g. tissue/air/bone interfaces at the head boundary), see Figure 3.3a. These large ripples have low spatial frequency components. However, the small anatomic structures of interest (e.g. microvessels) give rise to field variations with high spatial frequency components. This suggests a high pass filtering operation should be effective in separating the two kinds of field variations, making the brain matter easier to observe. For this project a 3D highpass Gaussian filter was constructed and applied to the raw phase images (as seen in Figure 3.3a), resulting in a cleaner phase image (Figure 3.3b), allowing the brain matter to be more observable.

Nevertheless, the phase image in Figure 3.3b still requires further processing before it is ready for the feature extraction step, with the problem being the background phase noise that lies outside the head. The pixels values representing phase noise with no signal (such as regions outside the head) are uncorrelated, where as pixels values representing signal with noise (such as regions inside the head) are more strongly correlated.
Figure 3.2 The outline of the steps involved in the preprocessing component of the pattern recognition system: (a) The raw phase image. (b) The result of applying a highpass filter to the raw phase image. (c) The result of apply an autocorrelation filter and then performing histogram equalization.
Figure 3.3 A typical raw phase image and its corresponding highpass filtered version:
(a) A raw phase image, notice the large ripples obscuring the brain matter. (b) The resulting highpass filter version of the raw phase image. Notice the result of applying a 3D highpass Gaussian filter allows the brain matter to be more observable.
Thus an autocorrelation filter was constructed to distinguish regions outside the head from regions inside the head, allowing the removal of the background noise from the phase image (Figure 3.4).

![Figure 3.4](image.png)

Figure 3.4 Plots of autocorrelation for various regions of a typical phase image: (a) The plot of the autocorrelation for a region in the air surrounding the head. Notice its resemblance to a Dirac delta function. (b) The plot of the autocorrelation for a region containing only brain tissue. (c) The plot of the autocorrelation for a region containing both blood vessels and suround brain tissue. (d) The plot of the autocorrelation for a region containing both brain tissue and air.
The autocorrelation filter was implemented as follows: for a given region in the phase image falling within the scanning window the autocorrelation values were computed, and sorted in decreasing order. The first and second (the largest and second largest) values were subtracted. If the difference was greater than the threshold (which was found empirically to be 30) that region had its pixel values set to -pi; otherwise the pixel values remain unaltered.

After applying the autocorrelation filtering process to the highpass filtered phase image the resulting phase image was histogram equalized to expand the range of weak phase differences due to small anatomical structures (e.g. blood vessels). This expansion of values further improves the image contrast. The histogram equalized phase image was then requantized from floating point numbers to integers ranging from 0 to 31. This was done to prepare the image for the subsequent feature extraction process, in particular the Haralick feature extractor (discussed in Section 3.3), whose measures are based on the co-occurrence matrix, and the analyzed image can only have integer values. Also the size of the co-occurrence matrix (and thus the computational cost) increases with the number of gray levels. Thirty two gray levels (0 to 31) was found empirically as a balance between proper representation of gray level variation and computational load (Figure 3.5). This completes the steps involved in the preprocessing component. The resulting phase image is now ready for the feature extraction component, which is discussed in the next section.
Figure 3.5  The resulting phase image after it has been highpass filtered, ran through the autocorrelation filtering process, been histogram equalized, and finally having its values requantized to integers ranging from 0 to 31.

3.3 Feature Extraction: Phase Image Texture Features

The feature extraction process is used to characterize an object to be recognized by measurements whose values are very similar for objects in a particular category, and very different for objects in different categories. Automatic/computerized feature extraction of 7T MR data sets/images for the purpose of identifying blood vessel is made onerous by the following conditions. First, the nonuniform illumination due to inhomogeneous antenna sensitivity makes approaches based directly on voxel/pixel gray level less ineffective (e.g. thresholding techniques)  Second, the actual diameters of the microvessels being identified can be on the order of a voxel/pixel size or even smaller.  Third, the image contrast is low (though improved by the histogram equalization step employed during preprocessing).  The traditional methods are ineffective because they
fail to identify/capture the anisotropic properties of microvessels. As mentioned in Chapter 2, ideally blood vessels can be modeled approximately as circular cylinders, which are anisotropic geometric structures.

For this research project the chosen feature extraction methodology is anisotropic texture analysis. Here we define textures as an attribute representing the statistical spatial arrangement of gray levels of the pixels in a region of an image (Castleman 1996). The objective is to quantify the texture in an image so we can characterize local variation in gray levels within an object. The texture measurements are mapped into feature vectors to be used as input for the subsequent classification step. This section is divided into four subsections, each of which is dedicated to the investigation of finding useful texture features for input into the classifier, so that accurate identification of microvessels in selected regions/zones of interest within the human brain is feasible. For this project the voxel sizes of the data are 0.45 x 0.45 x 2.5 mm, hence the through-plane resolution is more than five times the in-plane resolution. Due to the limited information in the through plane direction (i.e. the 3rd dimension) relative to the in-plane directions (i.e. the 1st and 2nd directions), the features of interest used for this project are two dimensional.
3.3.1 Phase Image Marginal Statistics

An approach used frequently for texture analysis is based on statistical properties of the intensity histogram. For this study, we can apply this approach just as easily to phase images. One class of such measures is based on statistical moments of intensity values. Table 3.1 lists some common descriptors based on statistical moments and also on uniformity and entropy. These measures of texture are computed using only marginal histograms, thus carrying no information regarding the relative position of pixels with respect to each other hence no directionality. However, they still represent/measure some localized texture content by generating an intensity histogram as a scanning-window is raster scanned through the whole image, computing a texture feature from Table 3.1, and assigning the value to the center pixel of the scanning-window.
<table>
<thead>
<tr>
<th><strong>Mean</strong></th>
<th>$m = \sum_{i=0}^{L-1} z_i p(z_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>$\sigma = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)$</td>
</tr>
<tr>
<td><strong>Smoothness</strong></td>
<td>$R = 1 - 1/(1 + \sigma^2)$</td>
</tr>
<tr>
<td><strong>Third Moment</strong></td>
<td>$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$</td>
</tr>
<tr>
<td><strong>Uniformity</strong></td>
<td>$U = \sum_{i=0}^{L-1} p^2(z_i)$</td>
</tr>
<tr>
<td><strong>Entropy</strong></td>
<td>$e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$</td>
</tr>
</tbody>
</table>

Table 3.1 Descriptors of texture based on intensity histogram, where $z$ is a random variable indicating the phase value, $p(z_i)$ is the histogram of the phase image values in a region, and $L$ is the number of possible gray levels in the image. Note these statistics cannot sense directional differences.
3.3.2 Marginal Statistics of Phase Image Spatial Derivatives

Estimating the anisotropy and the corresponding orientation fields of images or zones within images has a wide range of applications in texture analysis. From fingerprints to transmission electron microscopy, many cases require the identification of anisotropy as a major characteristic of texture or pattern recognition. As a result, numerous techniques for estimating local orientation fields have been proposed (Bergonnier, Hild and Roux 2007). Gradients which determine rates of change relative to position provide some texture information and directional sensitivity. Rao introduced one of the earliest approaches to determine the principal orientation field from an image by using the direction of local gray-level gradients (Rao 1990). Rao’s approach was based on intensity/magnitude images; however, as with the marginal statistics in Section 3.3.1, we can apply Rao’s method to phase images. The local principal anisotropy direction is perpendicular to the gradient. Let $f(x,y)$ represent a gray-scale phase image, and $\nabla f$ its gradient. Based on the discrete (pixel based) field, a approximate $G$ of the gradient $\nabla f$ is computed and denoted as

$$ G = \begin{bmatrix} D_x(x,y) \\ D_y(x,y) \end{bmatrix} $$  \hspace{1cm} (Equation 3.1)

where $G$ is a two element vector with x-component and y-component of the gradient and $D_x(x,y) = \partial f/\partial x$ and $D_y(x,y) = \partial f/\partial y$. 

45
Rao’s method is to compute over window $W$ the mean tensor product

$$\Gamma = \frac{\Sigma_W G^T G}{N_w}$$  \hspace{1cm} (Equation 3.2)

where $(.)^T$ denotes the transpose and $N_w$ is the number of pixels in the scanning-window $W$. The window $W$ is raster scanned throughout the image and at each location a $\Gamma$ matrix is computed. The $\Gamma$ matrix is then diagonalized using singular value decomposition, resulting in the computation of the eigenvalues and eigenvectors of $\Gamma$. The direction of the anisotropy axis is then perpendicular to the eigenvector associated with the larger of the two eigenvalues. The local strength of the direction field that corresponds to an anisotropy degree can be defined as

$$\text{Str} = \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2}}$$  \hspace{1cm} (Equation 3.3)

where $\lambda_1$ and $\lambda_2$ correspond to the large and small eigenvalues of Equation 3.2. Thus each pixel in the phase image can be associated with a pair of eigenvalues, which can encode local anisotropic strength/information. In this case the eigenvalue(s) or their arithmetic manipulations serve as the feature(s) of interest.
3.3.3 Second Order Spatial Partial Derivatives of the Phase Image

Taking spatial derivatives of the gradient components, we can use Hessian matrix elements as further measures of anisotropic texture. A Hessian is a symmetric matrix that describes the local curvature of a function with multiple variables. In this approach, \( D_{xx}(x,y) \), \( D_{yx}(x,y) \), \( D_{xy}(x,y) \), and \( D_{yy}(x,y) \) are computed and defined in Table 3.2. A window \( W \) is raster scanned through the phase image \( f(x,y) \), a matrix is generated for each pixel contained in \( W \), and averaged together, similar to what is done in Equation 3.2. The small and large eigenvalues of the resulting matrix \( H \) are taken to be the features of interest, where

\[
H = \frac{\sum_{w} \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}}{N_w} \quad (\text{Equation 3.4})
\]
Table 3.2 1st and 2nd order partial spatial derivatives.

One can view the $H$ matrix as the mean of a 2x2 Hessian matrix, and as the $H$ matrix is computed so are its eigenvalues, $\lambda_1$ and $\lambda_2$, where we take $\lambda_1 \geq \lambda_2$. The eigenvalues measure the convexity and concavity the corresponding eigendirections. For a 7T phase image, where the vessels are darker than the background one should expect the eigenvalues of the Hessian matrix to behave such that $\lambda_1 > 0$ and $\lambda_2 \approx 0$ for vessel pixels, since the blood vessel would form local minima with respect to the surrounding brain tissue (Martinez-Perez, Hughes and Thom 2007).
3.3.4  Co-occurrence Texture Measures on Phase Image

Another widely used directionally sensitive set of texture measures was developed by Haralick, et al. These are based on the local joint statistics with the scanning-window (Haralick, Shanmugam and Dinstein 1973). The basis for these features is the gray-level co-occurrence matrix (GLCM), which are joint histograms of two random variables (Equation 3.5).

\[
\text{GCLM} = \begin{bmatrix}
    p(1, 1) & p(1, 2) & \cdots & p(1, N_g) \\
    p(2, 1) & p(2, 2) & \cdots & p(2, N_g) \\
    \vdots & \vdots & \ddots & \vdots \\
    p(N_g, 1) & p(N_g, 2) & \cdots & p(N_g, N_g)
\end{bmatrix}
\]  

(Equation 3.5)

This matrix is square with dimension \( N_g \), where \( N_g \) is the number of gray levels in the image. Element \([i, j]\) of the matrix is generated by counting the number of times a pixel with value \( i \) is adjacent to a pixel with value \( j \) and then dividing the entire matrix by the total number of such comparisons made. Each entry therefore represents the probability that a pixel with value \( i \) will be found adjacent to a pixel of value \( j \). Figure 3.6 illustrates how a GCLM can be calculated for its first three values.
Figure 3.6 A given input with 8 gray levels (1 through 8) and its corresponding GCLM, defined for the nearest neighbor in the horizontal direction.

The GLCM’s (1,1) element contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. The GCLM (1,2) element contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. One can continue processing the given input image in a similar fashion, scanning the image for other pixel pairs \((i,j)\) and recording the sums in the corresponding elements of the GLCM (Matlab 2010).

Adjacency can be defined to occur in each of the four directions in 2D, namely, horizontal, vertical, left, and right diagonals – see Figure 3.7. Accompanying the direction specification one must also specify the distance (one pixel-length, two pixel-length, etc.), along with the size of the scanning-window (3x3, 5x5, etc.) which
determines the spacing and scale, respectively, one wishes to analyze the image’s texture. Thus GLCMs can be calculated for each of the four directions, for a specified distance within a given scanning-window size.

![Figure 3.7](image.png)

Figure 3.7 Four directions of adjacency as defined for calculation of the Haralick texture features: (a) The horizontal direction. (b) The vertical direction. (c) The left diagonal direction. (d) The right diagonal direction.

Haralick also described 14 statistics that can be defined from the GLCM with the intent of describing the texture of the image (or local regions within the image), see Table 3.3 (Haralick, Shanmugam and Dinstein 1973). Since the co-occurrence matrix is inherently directional and scale dependent, the derived texture features are capable of quantifying the local anisotropic properties within the various regions of the phase image. For a given texture feature \( F_i \), local anisotropy can be indicated as its value along a particular direction differing from its value along a different direction. For an example, \( x \)-direction (horizontal) and \( y \)-direction (vertical) GCLMs can be computed, then for each pixel location we can assign the ordered pair \((F_{ix}, F_{iy})\) that represents the feature components from the two corresponding GCLM matrices. Thus for each pixel location there is an associated ordered pair representing the texture’s value at the location in each
of the specified directions. This can allow one to examine the behavior the texture feature along the x and y directions in a given pixel neighborhood within the phase image, by pixelwise comparison of $F_{ix}$ and $F_{iy}$.

Table 3.3  The 14 Haralick Features

<table>
<thead>
<tr>
<th>Angular Second Moment</th>
<th>$\sum_{i,j} p(i,j)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>$\sum_{n=0}^{N_x-1} n^2 { \sum_{i=1}^{N_y} \sum_{j=1}^{N_x} p(i,j) }$, $</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\sum_{n=0}^{N_x-1} \sum_{i,j} (p(i,j) - \mu_x \mu_y)^2 \sigma_x \sigma_y$</td>
</tr>
<tr>
<td></td>
<td>where $\mu_x$, $\mu_y$, $\sigma_x$, and $\sigma_y$ are the means and std. deviations of $p_x$ and $p_y$, the partial probability density functions</td>
</tr>
<tr>
<td>Sum of Squares: Variance</td>
<td>$\sum_{i,j} (i - \mu)^2 p(i,j)$</td>
</tr>
<tr>
<td>Inverse Difference Moment</td>
<td>$\sum_{i,j} \frac{i}{</td>
</tr>
<tr>
<td>Sum Average</td>
<td>$\sum_{i=1}^{2N_x} \sum_{j=1}^{2N_y} p_{x+y}(i)$</td>
</tr>
<tr>
<td></td>
<td>where $x$ and $y$ are the coordinates (row and column) of an entry in the co-occurrence matrix, and $p_{x+y}(i)$ is the probability of co-occurrence matrix coordinates summing to $x+y$</td>
</tr>
<tr>
<td>Sum Variance</td>
<td>$\sum_{i=1}^{2N_x} (i - f_8)^2 p_{x+y}(i)$</td>
</tr>
<tr>
<td>Sum Entropy</td>
<td>$- \sum_{i=1}^{2N_x} p_{x+y}(i) \log { p_{x+y}(i) } = f_8$</td>
</tr>
<tr>
<td>Entropy</td>
<td>$- \sum_{i,j} \log { p(i,j) }$</td>
</tr>
<tr>
<td>Difference Variance</td>
<td>$\sum_{i=0}^{N_x-1} i^2 p_{x-y}(i)$</td>
</tr>
<tr>
<td>Difference Entropy</td>
<td>$- \sum_{i=0}^{N_y-1} p_{x-y}(i) \log { p_{x-y}(i) }$</td>
</tr>
<tr>
<td>Info. Measure of Correlation 1</td>
<td>$\frac{HXY-HXY^2}{\max(HX,HY)}$</td>
</tr>
<tr>
<td>Info. Measure of Correlation 2</td>
<td>$(1 - \exp[-2(HXY-HXY)])$</td>
</tr>
<tr>
<td>Info. Measure of Correlation 2</td>
<td>where $HXY = - \sum_{i,j} p(i,j) \log { p(i,j) }$, $HX$, $HY$ are the entropies of $p_x$ and $p_y$, $HXY^1 = - \sum_{i,j} p(i,j) \log { p_x(i) p_y(j) }$, $HXY^2 = - \sum_{i,j} p_x(i) p_y(j) \log { p_x(i) p_y(j) }$</td>
</tr>
<tr>
<td>Max. Correlation Coeff.</td>
<td>Square root of the second largest eigenvalue of $Q$</td>
</tr>
<tr>
<td></td>
<td>where $Q(i,j) = \sum_{k} p_k(i) p_k(j)$</td>
</tr>
</tbody>
</table>

52
The different types of local texture features that are calculated for the phase images in this research project have been presented. In summary there are the phase image marginal statistics (Section 3.3.1) that measures image texture using only marginal histograms providing no directionality information. Next there are the marginal statistics of phase image spatial derivatives (Section 3.3.2 and 3.3.3) that measure image texture by determining rates of change relative to position, thus providing directionally sensitive features. Finally there are the co-occurrence texture measures on the phase images (Section 3.3.4) that are based on local joint statistics within a scanning-window, and are sensitive to direction, spacing and scaling. Now that the feature extraction methodology has been presented, the next step in the pattern recognition process is classification.

3.4 Classification

As mentioned earlier, the task of the classifier component is to use the feature vectors provided by the feature extractor to assign the object to a category. The degree of difficulty of the classification problem depends on the variability in the feature values for objects in the same category relative to the difference between feature values for objects in different categories. The variability of feature values for objects in the same category may be due to complexity, and may be due to noise. Noise can be defined as any property of the sensed pattern which is not due to the true underlying model but instead due to randomness in the sensors composition (e.g. 1/f noise). All nontrival pattern recognition problems involve noise in some form. As such, perfect classification is often impossible, and a more practical task is to determine the probability that the observed
object/pattern belongs to one of the categories (Bishop 1995). Thus the most natural framework in which to formulate solutions of pattern recognition problems is a statistical one, which recognizes the probabilistic nature of both the information we seek to process, and the form in which we should express our results. Along with specifying or designing a particular classifier, one must also specify the method for evaluating its performance.

In recent years neural networks have found many applications in such diverse fields as modeling, time series analysis, signal processing, control, and pattern recognition. Neural networks are composed of simple elements (neurons) operating in parallel; see Figure 3.8 for the model of a neuron (Haykin 1998). These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. One can train a neural network to perform a particular function by adjusting the values of the connections (called synaptic weights) between elements.
Figure 3.8 Signal flow graph model of a neuron. $x_0$ through $x_n$ represent the inputs, $w_0$ through $w_n$ represent their associated synaptic weights, $\Phi(\bullet)$ represents the transfer function, and $y$ represents the neuron’s output. Note the input $x_0 = +1$ and the associated synaptic weight $w_0$ is equal to the bias of the neuron.

Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output, as illustrated in Figure 3.9. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are needed to train a network. The neuron model and the architecture of a neural network describe how a network transforms its inputs into an output (Haykin 1998). The manner in which the neurons of a neural network are structured (or connected) is intimately linked with the learning algorithm used to train the network (i.e. adjusting the synaptic weights). The behavior of the network depends on both the synaptic weights and the input-output function (transfer function) that is
specified for the units. This function typically falls into one of three categories: linear, threshold, or sigmoid (Haykin 1998). There are various network structures (also called network architectures) in use, the present discussion focuses on just one, namely, multilayer (ML) feedforward networks (Figure 3.10).

Figure 3.9 A block diagram representation of a typical neural network.
ML feedforward networks can, at least in principle, provide the optimal solution to an arbitrary classification problem. They implement linear discriminants, but in a space where the inputs have been mapped nonlinearly. The key power provided by such networks is that they admit simple algorithms where the form of the nonlinearity can be learned from training data, and one of the most popular methods for training ML networks is based on gradient descent in error – the backpropagation algorithm – which is very useful, but also relatively simple (Haykin 1998). From a design standpoint, once a ML network is chosen (sometimes along with its learning algorithm and corresponding transfer functions) the next step is to choose the number of inputs nodes, number of layers, the number of neurons in each layer. These decisions are often made empirically, as was done for this research project and discussed later (Section 3.4.2).

Figure 3.10 Fully-connected feedforward network with two hidden layers and one output layer.
ML networks provide a powerful and speedy tool for building classifiers. In general, neural networks can be viewed as an extension of conventional techniques in statistical pattern recognition, and under certain circumstances the outputs of a neural network can be interpreted as probabilities (Bishop 1995). For this research project the objective is to use a binary classifier to make a decision on a pixel by pixel basis as to whether a given pixel belongs to a blood vessel or its surrounding tissue. Receiver operating characteristic (ROC) curves have been used extensively in many fields (e.g. signal detection theory, radiology, machine learning, etc) as a tool for evaluating the performance of binary classifiers (Fawcett 2006). Therefore this study uses a multilayer feed-forward neural network architecture as the classifier, and its ROC curve(s) as the method for evaluating its performance.

For this study one of the first problems to consider is the high dimensionality of the data been collected and analyzed. Even though feature extraction was used to help alleviate such problems (refer to Section 3.3) there remain a large number of possible feature combinations. For example, there are 14 Haralick features, four primary directions (0 degrees, 90 degrees, 45 degrees, 135 degrees) computed for various distances, and then there are 6 marginal phase image statistics, as well as 2 sets of eigenvalues pairs for the spatial derivative features. To further complicate the selection of features, multiple scanning-window sizes have been structured for each of these. Put all together there are hundreds of features as possible inputs. The remainder of this section discusses the systematic approach used to choose the best features (Section 3.4.1) for the chosen neural network architecture (Section 3.4.2).
3.4.1 Feature Selection Process

The main objective of a feature selection process is to reduce the number of features without losing the important/main information. One benefit of decreasing the number of inputs to the classifier (in this case a neural network) is that it reduces the training time and effort of measuring a number of features. Furthermore, beyond a certain point, adding features can actually lead to a reduction in the performance of the classification system (Bishop 1995). Nevertheless, when applying any feature reduction method, it is essential to ensure that the reduced features can still effectively map the desired neural network output. This research employs the forward stepwise discriminant feature selection process combined with histogram analysis and the use of the area under the ROC curve (called the ROC area index) as a measure of performance, as outlined in Figure 3.11.

The first step in the process involves computing a set of features using pre-specified parameters within a given feature extraction methodology, for example computing all Haralick texture features for a specified distance, direction, and scanning-window size. For a preliminary examination of each feature’s potential to distinguish pixels belonging to blood vessels from pixels belonging to the surrounding tissue, a histogram is formed to examine the overlap between feature values corresponding to vessel pixels versus nonvessel pixels. As stated earlier, the degree of the difficulty of the classification problem depends on the variability in the feature values for objects in the same category relative to the difference between feature values for objects in different categories. Thus the best features are those whose histogram values corresponding to
blood vessel are more distinct from those corresponding to brain tissue. In other words, the best features produce the most separation (or least overlap) of the distributions of the two classes in the histogram (Figure 3.12).

Figure 3.11 Flow Chart outlining the feature selection process.
Figure 3.12 A plot of the histograms of a feature variable $x_1$ and $x_2$: (a) An example of a feature variable ($x_1$) with relatively good discrimination power. For this feature the nonvessel pixels tend to be centered around zero then rapidly decrease in concentration on either side of zero, and extending pass ten. The blood vessel pixels are centered around two, and remains relatively concentrated up to about five and extents all the way pass twenty. (b) An example of a feature variable ($x_2$) with relatively poor discrimination power. Both nonvessel and vessel pixels values are grouped together, and the two distributions almost completely overlap.
In Figure 3.12, $x_1$ and $x_2$ are two candidate features for inputs to the neural network shown in Figure 3.10. In Figure 3.12a the nonvessel and vessel pixel values are centered at two different locations and are visually distinct in their relative concentrations and range of values. Conversely, in Figure 3.12b one can quickly notice this particular feature variable has poor discriminatory power, because both pixel value distributions almost completely overlap. By using this simple visual inspection technique one can quickly determine $x_1$ is worth further consideration and $x_2$ should be eliminated. This visual histogram inspection technique is employed repeatedly, for various feature variables produced using the aforementioned feature extraction methodologies to help make a quick decision on which features (and to an extent which methodology) can move forward for further consideration, and which should be eliminated.

Once the histogram inspection step is completed the remaining feature variable candidates are subjected to the next step. In the second step, we reduce the amount redundant information and increase the efficiency of the forward stepwise discriminant feature selection process by combining a measure of each feature’s individual discriminate power with correlation analysis. This process is discussed in detail in Section 4.2. The third step consists of the forward stepwise discriminant selection process. The process works has follows: 1) the best performing feature of step 2 is paired with every other feature, and feed into the network. Then the best performing pair is recorded, 2) each of the remaining features is added to the best performing pair (forming a triplet). Then the best performing triplet is recorded. The process continues until the ROC area index ceases to increase significantly with the addition of new
features. The next section discusses the general neural network architecture used in this research project.

3.4.2 Assignment of the Neural Network Architecture

As mentioned earlier, for this study a ML feedforward network is the architecture of choice and was implemented using Matlab software, see Figure 3.13. The architecture includes one hidden layer of neurons, and a single neuron for the output layer. The network consists of tan-sigmoid transfer functions in both the hidden layer and the output layer, and uses the Scaled-Conjugate Gradient algorithm for training. Figure 3.13 illustrates the general form of the architecture. During the feature selection process the number input layer nodes (i.e. the number of features used as input) was varied along with the number of hidden neurons in an effort to simultaneously find the optimal combination features and hidden layer neurons, using the ROC area index as the figure-of-merit. In the next chapter the final of form the ML network architecture and its optimal input features are presented.
Figure 3.13 The general form of the fully-connected feedforward architecture used for this research project.
Chapter 4

Results and Discussion

For this study the data consist of 7T phase images of the human brain from three subjects from which five slices were selected from each. The slices from each subject were centered on the corpus callosum region of the brain. The corpus callosum region was chosen for its anatomical distinction and clinical significance, because it is frequently involved by infiltrating tumors of the brain, such as gliomas, sarcomas, and occasional metastatic tumors (Ucmakli 1972). Each subject was scanned twice; first the subject was scanned, then given at dose of 250mg of caffeine waited ten minutes and scanned again. Hence there are six distinct sets of data, a caffeinated and uncaffeinated image set for each of the three subjects.

The caffeine dose causes the blood vessels to constrict which slows the flow of blood, in particular oxygenated blood from the arteries. Since the oxygen consumption in the rest of the body remains the same, the vascular constriction leads to a deeper depletion of oxygen causing a higher percentage/fraction of deoxyhemoglobin in the
veins. The increased presence of deoxyhemoglobin causes stronger paramagnetic susceptibility effects leading to more phase change in the local spins. Since the increase in phase change occurs in the same vessels, we can find/observe veins that were barely (or undetectable) by a human observer before the caffeine dose. Therefore in effect, the caffeine via increased doxyhemoglobin concentration acts as a contrast agent.

As a result of the separate caffeinated and uncaffeinated image sets, we can attempt to develop algorithms to automatically detect blood vessels in the uncaffeinated phase images that are not specifically resolvable, and test the effectiveness of the algorithm using the corresponding caffeinated phase images. Chapter 2 presented the underlying physical theory of the paramagnetic susceptibility effects that are represented in the phase images. Chapter 3 introduced the mathematical principles behind the texture features that attempt to distinguish blood vessel from the surrounding tissues. This chapter provides an evaluation of the features’ and classifier’s ability to identify blood vessels within the phase image data set.

As mentioned above there are two broad categories for the image data sets, labeled Caffeinated and Uncaffeinated data. Each data set consists of 5 slices from each of the three subjects. Within each slice 5 regions-of-interest (ROI) of size 20X20 pixels were selected and their identifiable blood vessels were hand marked by a radiologist using the MIPAV software. For each subject the uncaffeinated images were marked-up first without referring to the caffeinated images, and then the caffeinated images were highlighted. Once the mark-ups were completed for all three subjects, subject 1’s data was reserved for evaluating the feature selection process and for training the neural
network; as well as for performing preliminary testing. Subjects 2 and 3 data were used for more formal testing. The rest of this chapter is dedicated to discussing the results of the investigative process used to find the features, and then testing and training the classifier. In Section 4.1, we present the results of the histogram analysis and the initial feature candidates from each of the feature extraction methodologies discussed in Chapter 3. In Section 4.2, the outcomes of the correlation analysis performed on the initial features, and the list of the least correlated and best performing features (termed neural network input candidates) are presented. In Section 4.3, the features’ and classifier’s preliminary performances are evaluated via the ROC area index. Section 4.4 complements Section 4.3 with a more qualitative assessment via color image display. Finally in Section 4.5, the results of formally testing the classifier and chosen features on subjects 2 and 3 are presented.

4.1 Histogram Analysis

Within this section histogram analysis is used as simple yet effective visual inspection tool to assess the performance of the features extracted using the various methodologies presented in Chapter 3.

4.1.1 Histogram Analyses of Phase Image Marginal Statistics

The individual marginal texture measures were defined in Table 3.1 of the previous chapter and are relisted here in Table 4.1 for convenience. As mentioned in Chapter 3, marginal texture measures are computed using only marginal histograms; thus
carrying no information regarding the relative position of pixels with respect to each other; hence, they are not directionally sensitive. However, they still represent/measure localized texture content by generating a histogram of phase values as a scanning-window is moved through the whole image. For this study, 4 scanning-window sizes (W) were used, namely, 3x3, 5x5, 7x7, and 9x9. For each of the window sizes, the best performing feature(s) were selected as possible neural input candidates. Their performance was judged visually by the amount of separation created in their corresponding histograms (i.e. their ability to take on distinct values for pixels belonging to vessels from pixels belonging to the surrounding tissue). For the case of the 3x3 window, the mean and standard deviation were the two best features (see Figure 4.1). As expected from observing the phase images, the blood vessels tend to be localized in regions that are darker (i.e. lower pixel values) than the surrounding tissue, and tend to spread out slightly over a larger range of values, as indicated in Figure 4.1.
<table>
<thead>
<tr>
<th><strong>Mean</strong></th>
<th>[ m = \sum_{i=0}^{L-1} z_i p(z_i) ]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>[ \sigma = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i) ]</td>
</tr>
<tr>
<td><strong>Smoothness</strong></td>
<td>[ R = 1 - 1/(1 + \sigma^2) ]</td>
</tr>
<tr>
<td><strong>Third Moment</strong></td>
<td>[ \mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) ]</td>
</tr>
<tr>
<td><strong>Uniformity</strong></td>
<td>[ U = \sum_{i=0}^{L-1} p^2(z_i) ]</td>
</tr>
<tr>
<td><strong>Entropy</strong></td>
<td>[ e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) ]</td>
</tr>
</tbody>
</table>

Table 4.1  Descriptors of texture based on intensity histogram, where \( z \) is a random variable indicating the phase value, \( p(z_i) \) is the histogram of the phase image values in a region, and \( L \) is the number of possible gray levels in the image. Note these statistics cannot sense directional differences.
Figure 4.1 Histogram of the candidate marginal statistics for W = 3x3: (a) The mean. (b) The standard deviation.

For W = 5x5, the mean, standard deviation, entropy, and uniformity are the chosen as input candidates (see Figure 4.2). For W = 7x7, the best performing features were uniformity and entropy (see Figure 4.3). For W = 9x9, the features of interest were uniformity and entropy (see Figure 4.4). The uniformity feature attempts to characterize the intuitive understanding when vessel pixels are in the scanning window, there is less uniformity of pixel values than when there is no vessels in the window. Entropy is a
measure of randomness or the uncertainty associated with pixel values in a given region. The entropy feature supports our intuition that localized regions containing blood vessels would possess more randomness than a region containing only tissue. For some of the above features there is still quite a bit of overlap between vessels and nonvessels in their histograms, but they remain in the pool as having possible discriminatory power when used in combination with the other features. Refer to Table 4.2 for a list of potential candidates from the marginal statistic feature extraction methodology remaining in the pool after histogram analysis.
Figure 4.2 Histogram of the candidate marginal statistics for $W = 5x5$: (a) The mean. (b) The standard deviation (c) The uniformity (d) The entropy
Figure 4.3 Histograms of candidate marginal statistics for $W = 7 \times 7$: (a) The uniformity. (b) The entropy.
Figure 4.4  Histograms for the candidate marginal statistics for $W = 9 \times 9$ (a) The uniformity. (b) The entropy.
### Features

<table>
<thead>
<tr>
<th>W</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>3x3</td>
<td>Mean and Standard Deviation</td>
</tr>
<tr>
<td>5x5</td>
<td>Mean, Standard Deviation, Uniformity, and Entropy</td>
</tr>
<tr>
<td>7x7</td>
<td>Uniformity and Entropy</td>
</tr>
<tr>
<td>9x9</td>
<td>Uniformity and Entropy</td>
</tr>
</tbody>
</table>

Table 4.2 Potential feature candidates of the marginal statistic feature extraction methodology.

### 4.1.2 Histogram Analyses of Statistics of Phase Image Spatial Derivatives

The statistics for the phase image spatial derivatives were discussed in Chapter 3. The two key matrices the $\Gamma$ matrix and the $H$ matrix, which for convenience are provide here.

$$\Gamma = \frac{\sum W G^T G}{N_w} \quad \text{(Equation 4.1)}$$

And

$$H = \frac{\sum W \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}}{N_w} \quad \text{(Equation 4.2)}$$
Where

\[ G = \begin{bmatrix} D_x(x,y) \\ D_y(x,y) \end{bmatrix} \]  
(Equation 4.3)

and where

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_x )</td>
<td>( \frac{\partial}{\partial x} )</td>
</tr>
<tr>
<td>( D_{xx} )</td>
<td>( \frac{\partial^2}{\partial x^2} )</td>
</tr>
<tr>
<td>( D_y )</td>
<td>( \frac{\partial}{\partial y} )</td>
</tr>
<tr>
<td>( D_{yy} )</td>
<td>( \frac{\partial^2}{\partial y^2} )</td>
</tr>
<tr>
<td>( D_{xy} = D_{yx} = \frac{\partial}{\partial x \partial y} = \frac{\partial}{\partial y \partial x} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 1st and 2nd order partial spatial derivatives.

The eigenvalues of the two matrices that are of primary concern. Similar to the previous section, various scanning-window sizes were used, namely, W= 3x3, 5x5, 7x7, and 9x9 for the Rao based approach, and sizes of 1x1, 3x3 5x5, 7x7, and 9x9 for Hessian matrix approach. For each, a window W was scanned through the image and at each pixel location a matrix \( I \) or \( H \) was computed and its eigenvalues determined. Features of
interest in both case were either the eigenvalues themselves or their arithmetically
manipulated version, namely, the primary eigenvalues, the secondary eigenvalues, the
normalized primary eigenvalue and the anisotropic degree (Str), defined by equations 4.4
and 4.5, respectively.

\[
\text{Normalized Primary Eigenvalue} = \frac{\lambda_1}{\sqrt{\lambda_1^2 + \lambda_2^2}} \quad \text{(Equation 4.4)}
\]

\[
\text{Str} = \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2}} \quad \text{(Equation 4.5)}
\]

Implementation of the Rao based method of diagonalizing the \(\Gamma\) matrix and find its
eigenvalues yielded no features of interest for any of the window sizes because their
histograms were severely overlapped. The eigenvalues resulting from the diagonalization
of the \(H\) matrix yielded better results. For case \(W = 1x1\) the normalized primary
eigenvalue and the primary eigenvalue itself were the two best features. For the case of
\(W = 3x3\) the normalized primary eigenvalue and primary eigenvalue were the best two
features. For \(W = 5x5, 7x7, \) or \(9x9\) the eigenvalues provided no features of interest, and
such an outcome is to be expected. The \(H\) matrix describes the local curvature in the
phase image, so it makes sense that as \(W\) size increases there would be a point at which
the \(H\) matrix fails to accurately describe/represent the local curvature sufficiently enough
to distinguish blood vessels from tissue. Refer to Table 4.3 for a list of potential candidates from the $H$ matrix feature extraction methodology after histogram analysis.

<table>
<thead>
<tr>
<th>W</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x1</td>
<td>Primary EigenValue and Normalized Primary EigenValue</td>
</tr>
<tr>
<td>3x3</td>
<td>Primary EigenValue and Normalized Primary EigenValue</td>
</tr>
</tbody>
</table>

Table 4.4 Potential feature candidates of the H matrix feature extraction methodology

4.1.3 Histogram Analyses of Co-occurrence Textures on Phase Image

The features of interest for this approach are the Haralick features, which are joint statistics derived from the GLCMs that were discussed in Chapter 3. The GCLM is computed along with its derived features as a window (W) is scanned through the image. For this study the GCLMs were specified to be sensitive to each of four directions, namely, horizontal, vertical, left diagonal, and right diagonal, and the distance for all cases was 1 pixel. The selected sizes for W were: 3x3, 5x5, 7x7, and 9x9. For each scanning-window size and specified direction, each of the Haralick feature were computed, as list in Table 3.8. Also included in the directional specification were combinations of directions termed horizontal-vertical combinations and right-left diagonal combination. The specifying the horizontal-vertical combinations allowed the
GCLM to be sensitive in both the horizontal and vertical directions. Essentially, it adds the horizontally specified GCLM to the vertically specified GCLM and then the Haralick features are computed using this combined GCLM, and similarly for the right-left diagonal combination GCLM. Hence, for a given scanning-window size there 6 directional specifications, namely, the horizontal, vertical, left diagonal, right diagonal, horizontal-vertical combination and right-left diagonal combination.

Refer to Table 4.5 for the complete list of co-occurrence texture feature candidates. The numbers listed in each entry of Table 4.5 correspond to the Haralick features such those illustrated in Table 4.6. For example, the entries in Table 4.5 occupying 2\textsuperscript{nd} row and 2\textsuperscript{nd} column (i.e. 6 and 7) correspond to Haralick features Sum Average and Sum Variance in Table 4.6. This marks the completion of histogram analysis, in the next section we begin to pare down the list of potential candidate features listed in Tables 4.2, 4.3, and 4.4.
<table>
<thead>
<tr>
<th>Directional Sensitivity</th>
<th>W = 3x3</th>
<th>W = 5x5</th>
<th>W = 7x7</th>
<th>W = 9x9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td>6,7</td>
<td>1,6,7,8,9,12,13</td>
<td>1,5,8,9,13</td>
<td>1,5,8,9,13</td>
</tr>
<tr>
<td>Vertical</td>
<td>6,7</td>
<td>1,5,6,7,8,9,11,13</td>
<td>1,2,5,8,9,11,13</td>
<td>1,5,8,9,11,13</td>
</tr>
<tr>
<td>Horiz. – Vert. Comb.</td>
<td>1,5,6,7,8,9</td>
<td>1,5,6,7,8,9,11,13</td>
<td>1,2,8,9,11,13</td>
<td>1,2,3,5,8,9,10,11,13</td>
</tr>
<tr>
<td>Right-Diagonal</td>
<td>6,7,13,12,9</td>
<td>1,5,6,7,8,9,12,13</td>
<td>1,5,8,9,11,13</td>
<td>1,8,9,11,13</td>
</tr>
<tr>
<td>Left-Diagonal</td>
<td>6,7,12</td>
<td>1,5,6,7,8,9,11,12,13</td>
<td>1,5,8,9,11,12,13</td>
<td>1,2,3,5,8,9,10,12,13</td>
</tr>
<tr>
<td>Right-Left Comb.</td>
<td>1,5,6,7,9,12,13</td>
<td>1,5,6,7,8,9,11,12,13</td>
<td>1,2,5,8,9,11,13</td>
<td>1,2,3,5,8,9,10,11,13</td>
</tr>
</tbody>
</table>

Table 4.5 Potential Haralick feature candidates
<table>
<thead>
<tr>
<th>Coded Numbers</th>
<th>Haralick Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Angular Second Moment</td>
</tr>
<tr>
<td>2</td>
<td>Contrast</td>
</tr>
<tr>
<td>3</td>
<td>Correlation</td>
</tr>
<tr>
<td>4</td>
<td>Sum of Squares: Variance</td>
</tr>
<tr>
<td>5</td>
<td>Inverse Difference Moment</td>
</tr>
<tr>
<td>6</td>
<td>Sum Average</td>
</tr>
<tr>
<td>7</td>
<td>Sum Variance</td>
</tr>
<tr>
<td>8</td>
<td>Sum Entropy</td>
</tr>
<tr>
<td>9</td>
<td>Entropy</td>
</tr>
<tr>
<td>10</td>
<td>Difference Variance</td>
</tr>
<tr>
<td>11</td>
<td>Difference Entropy</td>
</tr>
<tr>
<td>12</td>
<td>Information Measure of Correlation #1</td>
</tr>
<tr>
<td>13</td>
<td>Information Measure of Correlation #2</td>
</tr>
</tbody>
</table>

Table 4.6 The encoded numbers and their corresponding Haralick features used in Table 4.5
4.2 Correlation Analysis and Complete List of Candidates

After histogram analysis screening, we have total of 169 candidate features from Tables 4.2, 4.4 and 4.5. We reduce the amount redundant information and increase the efficiency of the forward stepwise discriminant feature selection process by combining a measure of each feature’s individual discriminate power with cross correlation analysis between features. A measure of each feature’s individual discriminate power is assessed by feeding each individual feature into a feedforward neural network consisting of a single neuron in the hidden layer and one neuron in the output layer, and computing the ROC area index. The neural network architecture is similar to the one discussed in Section 3.4.2, however implemented with a single neuron in the hidden layer. Using such a network to test individual inputs uses the nonlinearity of the neuron decision and the training algorithm to assess the best bias for each input.

For cross correlation assessment, a correlation coefficient matrix was computed for each set of features corresponding to Tables 4.2, 4.4, and 4.5, and the absolute value was taken for each entry in the matrix. Any feature correlating with another feature at or above a threshold of 0.88 was flagged, and features with the smaller ROC area index were eliminated. One last correlation coefficient matrix was computed for all remaining features. Once again the absolute values of the entries were calculated, and any features correlated above 0.88 were flagged. The flagged features with the smaller ROC area index were eliminated.

The process discussed above was fast and effective in removing features with relatively weak discriminate power and/or redundant information. Table 4.7 provides the
list of features to be used as candidate for the forward stepwise discriminant feature selection process. We have reduced the number of possible candidates from 169 to 33. During the forward feature selection process the feature with the highest ROC area index from the above mentioned process is paired with every other feature, and feed into the network with multiple neurons. Then the best performing pair is recorded. Each of the remaining features is added to the best performing pair (forming a triplet). Then the best performing triplet is recorded. The process continues until the ROC area index ceases to increase significantly with the addition of new features. Table 4.8 lists the input features after completing the forward feature selection process.
<table>
<thead>
<tr>
<th>Count</th>
<th>W</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1x1</td>
<td>Primary EigenValue of $\mathbf{H}$ matrix</td>
</tr>
<tr>
<td>2</td>
<td>1x1</td>
<td>Normalized Primary EigenValue of $\mathbf{H}$ matrix</td>
</tr>
<tr>
<td>3</td>
<td>3x3</td>
<td>Primary EigenValue of $\mathbf{H}$ matrix</td>
</tr>
<tr>
<td>4</td>
<td>3x3</td>
<td>Normalized Primary EigenValue of $\mathbf{H}$ matrix</td>
</tr>
<tr>
<td>5</td>
<td>3x3</td>
<td>Horizontal-Vertical combo. Inverse Difference Moment</td>
</tr>
<tr>
<td>6</td>
<td>3x3</td>
<td>Horizontal-Vertical combo. Entropy</td>
</tr>
<tr>
<td>7</td>
<td>5x5</td>
<td>Horizontal-Vertical combo. Inverse Difference Moment</td>
</tr>
<tr>
<td>8</td>
<td>5x5</td>
<td>Horizontal-Vertical combo. Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>9</td>
<td>7x7</td>
<td>Horizontal-Vertical combo. Angular Second Moment</td>
</tr>
<tr>
<td>10</td>
<td>7x7</td>
<td>Horizontal-Vertical combo. Contrast</td>
</tr>
<tr>
<td>11</td>
<td>5x5</td>
<td>Horizontal Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>12</td>
<td>9x9</td>
<td>Horizontal Inverse Difference Moment</td>
</tr>
<tr>
<td>13</td>
<td>5x5</td>
<td>Vertical Inverse Difference Moment</td>
</tr>
<tr>
<td>14</td>
<td>5x5</td>
<td>Vertical Difference Entropy</td>
</tr>
<tr>
<td>15</td>
<td>7x7</td>
<td>Vertical Contrast</td>
</tr>
<tr>
<td>16</td>
<td>9x9</td>
<td>Vertical Inverse Difference Moment</td>
</tr>
<tr>
<td>17</td>
<td>3x3</td>
<td>Right-Left Diagonal combo. Sum Variance</td>
</tr>
<tr>
<td>18</td>
<td>3x3</td>
<td>Right-Left Diagonal combo. Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>19</td>
<td>5x5</td>
<td>Right-Left Diagonal combo. Sum Entropy</td>
</tr>
<tr>
<td>20</td>
<td>5x5</td>
<td>Right-Left Diagonal combo. Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>21</td>
<td>7x7</td>
<td>Right-Left Diagonal combo. Difference Entropy</td>
</tr>
<tr>
<td>22</td>
<td>9x9</td>
<td>Right-Left Diagonal combo. Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>23</td>
<td>9x9</td>
<td>Right-Left Diagonal combo. Correlation</td>
</tr>
<tr>
<td>24</td>
<td>3x3</td>
<td>Right-Diagonal Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>25</td>
<td>3x3</td>
<td>Right-Diagonal Entropy</td>
</tr>
<tr>
<td>26</td>
<td>5x5</td>
<td>Right-Diagonal Info. Measure of Correlation 1</td>
</tr>
<tr>
<td>27</td>
<td>3x3</td>
<td>Left-Diagonal Info. Measure of Correlation 1</td>
</tr>
<tr>
<td>28</td>
<td>5x5</td>
<td>Left-Diagonal Difference Entropy</td>
</tr>
<tr>
<td>29</td>
<td>5x5</td>
<td>Left-Diagonal Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>30</td>
<td>7x7</td>
<td>Left-Diagonal Info. Measure of Correlation 2</td>
</tr>
<tr>
<td>31</td>
<td>9x9</td>
<td>Left-Diagonal Correlation</td>
</tr>
<tr>
<td>32</td>
<td>3x3</td>
<td>Marginal Statistics Standard Deviation</td>
</tr>
<tr>
<td>33</td>
<td>5x5</td>
<td>Marginal Statistics Standard Deviation</td>
</tr>
</tbody>
</table>

Table 4.7 List features to be used as candidates for the forward stepwise discriminant feature selection process.
Table 4.8 The feature candidates resulting from the forward feature selection process

These six features are judged to best characterize the texture within the phase images.

Figure 4.5 illustrates the classifier’s architecture and Table 4.9, Table 4.10, Table 4.11, and Table 4.12 provide the weights and biases following the network training. We label weight matrices connected to inputs input weights (IW) and label weight matrices coming from layer outputs layer weights (LW). Subscripts are used to identify the source node (second index) and the destination node (first index) for the various weights of network.

In the next section we begin our evaluation of the performance of the features and classifier.
Figure 4.5 The classifier’s architecture consists of 6 inputs, 12 hidden layer neurons, 1 output layer neuron. IW stands for input weights and LW stand for layer weights.
<table>
<thead>
<tr>
<th>Neurons</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.7685</td>
<td>0.9575</td>
<td>0.8837</td>
<td>0.0937</td>
<td>-1.1217</td>
<td>-0.0968</td>
</tr>
<tr>
<td>2</td>
<td>0.5011</td>
<td>1.0034</td>
<td>1.2179</td>
<td>-0.9474</td>
<td>1.0081</td>
<td>-0.6516</td>
</tr>
<tr>
<td>3</td>
<td>-0.1739</td>
<td>-0.9453</td>
<td>-0.9698</td>
<td>1.3524</td>
<td>-0.5848</td>
<td>0.5383</td>
</tr>
<tr>
<td>4</td>
<td>-0.5283</td>
<td>0.4237</td>
<td>0.4385</td>
<td>1.1116</td>
<td>-0.1987</td>
<td>0.9448</td>
</tr>
<tr>
<td>5</td>
<td>0.2710</td>
<td>0.7498</td>
<td>-1.1136</td>
<td>-1.1582</td>
<td>-0.4102</td>
<td>-0.7950</td>
</tr>
<tr>
<td>6</td>
<td>1.9827</td>
<td>0.2346</td>
<td>-0.4820</td>
<td>-0.2310</td>
<td>-0.4020</td>
<td>-0.6719</td>
</tr>
<tr>
<td>7</td>
<td>-0.3908</td>
<td>0.8152</td>
<td>-0.2783</td>
<td>1.2796</td>
<td>-0.8929</td>
<td>-0.5169</td>
</tr>
<tr>
<td>8</td>
<td>1.2682</td>
<td>-0.5142</td>
<td>0.4679</td>
<td>-1.2320</td>
<td>0.5313</td>
<td>0.3292</td>
</tr>
<tr>
<td>9</td>
<td>-1.3984</td>
<td>1.1528</td>
<td>0.8835</td>
<td>-0.6722</td>
<td>0.6267</td>
<td>-0.6855</td>
</tr>
<tr>
<td>10</td>
<td>-0.6074</td>
<td>0.6823</td>
<td>0.2032</td>
<td>0.6698</td>
<td>-1.2091</td>
<td>-1.2596</td>
</tr>
<tr>
<td>11</td>
<td>0.9393</td>
<td>0.1720</td>
<td>-0.9940</td>
<td>0.4978</td>
<td>-1.0016</td>
<td>-1.2981</td>
</tr>
<tr>
<td>12</td>
<td>-1.1854</td>
<td>1.4644</td>
<td>-0.1378</td>
<td>0.0107</td>
<td>0.3393</td>
<td>0.6198</td>
</tr>
</tbody>
</table>

Table 4.9 Input weight (IW) matrix. Each entry is the value of that connection’s weight from the input node to the corresponding neuron in the hidden layer.

<table>
<thead>
<tr>
<th>Neuron</th>
<th>Bias Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.2487</td>
</tr>
<tr>
<td>2</td>
<td>-1.5649</td>
</tr>
<tr>
<td>3</td>
<td>-1.4638</td>
</tr>
<tr>
<td>4</td>
<td>0.7653</td>
</tr>
<tr>
<td>5</td>
<td>0.6216</td>
</tr>
<tr>
<td>6</td>
<td>0.1321</td>
</tr>
<tr>
<td>7</td>
<td>0.5996</td>
</tr>
<tr>
<td>8</td>
<td>0.6718</td>
</tr>
<tr>
<td>9</td>
<td>-0.5930</td>
</tr>
<tr>
<td>10</td>
<td>1.9091</td>
</tr>
<tr>
<td>11</td>
<td>1.7182</td>
</tr>
<tr>
<td>12</td>
<td>-2.2857</td>
</tr>
</tbody>
</table>

Table 4.10 Hidden layer bias values for corresponding hidden layer neurons.
<table>
<thead>
<tr>
<th>Neuron</th>
<th>Weight Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3220</td>
</tr>
<tr>
<td>2</td>
<td>0.5784</td>
</tr>
<tr>
<td>3</td>
<td>-0.4981</td>
</tr>
<tr>
<td>4</td>
<td>-0.1902</td>
</tr>
<tr>
<td>5</td>
<td>0.8037</td>
</tr>
<tr>
<td>6</td>
<td>1.1083</td>
</tr>
<tr>
<td>7</td>
<td>0.1200</td>
</tr>
<tr>
<td>8</td>
<td>0.4208</td>
</tr>
<tr>
<td>9</td>
<td>-0.08287</td>
</tr>
<tr>
<td>10</td>
<td>-1.0455</td>
</tr>
<tr>
<td>11</td>
<td>-0.5891</td>
</tr>
<tr>
<td>12</td>
<td>-0.0526</td>
</tr>
</tbody>
</table>

Table 4.11 Output layer weights. Each entry is the value of that connection’s weight from the hidden layer neuron to the output neuron.

<table>
<thead>
<tr>
<th>Neuron</th>
<th>Bias Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0449</td>
</tr>
</tbody>
</table>

Table 4.12 Output layer bias value for the output neuron.
4.3 Performance Evaluation of Features and Classifier via ROC Area Index Using Training Data

The classifier was trained (i.e. weight and bias values set using the features in Table 4.8 as the inputs) using subject 1’s uncaffeinated image data. In particular, the training data consisted of three zones from each of the five slices. The features’ and classifier’s performance was initially tested using the remaining 2 zones from each of the five slices. The ROC curve and ROC area index was computed as a quantitative measure of performance (Figure 4.6). The ROC curve is generated by stepping through possible

![ROC Curve](image)

**Figure 4.6** The ROC curve for the subject 1 uncaffeinated ground truth image data. The ROC area index is 0.7901.
thresholds and calculating the true positive rate and the false positive rate. Actually use of the classifier requires a user-selected threshold for operation. For this dissertation research the threshold value is the Bayes’ point. The Bayes point is the point on the ROC curve that corresponds approximately to the knee of the curve.

The test also involved using the 2 zones from all five slices of the corresponding image data resulting from combining the caffeinated and uncaffeinated images. In particular, the caffeinated and uncaffeinated images had vessels that were highlighted manually by the radiologist and converted separately to binary images. An image union operation was performed on these two binary images, called combined ground truth image data. The combined ground truth was then used as the gold standard to test the features and trained classifier. Once again the ROC curve and ROC area index was computed as a quantitative measure of performance (Figure 4.7). Figure 4.8 illustrates a comparison of the classifier’s performance between using uncaffeinated ground truth and the combined ground truth as the gold standard for subject 1’s evaluation.
Figure 4.7 The ROC curve for subject 1 combined ground truth image data. The ROC area index is 0.7350.

Figure 4.8 Uncaffeinated ground truth Vs. Combined ground truth for subject 1.
There was a reduction in overall performance going from the uncaffeinated to the combined image data as judged by the ROC area index. However, there may be a corresponding reduction in the false positive rate going from the uncaffeinated to the combined ground truth image data. In other words, some of the false positives appearing in the uncaffeinated images may not be actual false positives, but vessels the radiologist initially missed when marking up the uncaffeinated images; however these were highlighted in the caffeinated images by the same radiologist. Since the combined images contain vessel information from both the uncaffeinated and caffeinated images, the classifier may have correctly classified vessels that were originally judged to be nonvessel based on only uncaffeinated data. This point is investigated in more detail the next section.

4.4 Performance Evaluation of Features and Classifier via Color-Coded Images Using Training Data

For a complementary and more qualitative assessment of features and classifiers performance we use color-coded RBG images. In particular, the binary image resulting from the classifier is assigned/coded as the green plane, the ground truth (i.e. radiologist highlighted images) is assigned/coded as the red plane, and the blue plane is zero. The three monochrome color planes/images are combined to form a RBG color image. Red corresponds to false negatives, that is, the pixels the radiologist judged to be vessel, but the classifier decided was brain tissue. Green corresponds to false positives, that is, the pixels the radiologist judge to be brain tissue, but the classifier decided was vessel.
Yellow results when the classifier and the radiologist agree the pixel is vessel. Gray or black pixels are where the classifier and radiologist agree there is nonvessel.

Figure 4.9 illustrates color coded image for subject 1 slice 3 using the uncaffeinated ground truth data. Figure 4.10 illustrates color coded image for the combined ground truth image data corresponding to Figure 4.9 using the Bayes point of subject 1 uncaffeinated images for reference. Figure 4.11 displays zoomed in version of the two previous images.

To reiterate, for all subjects 1, 2, and 3 the inputs into the classifier are the 6 texture features computing using only the uncaffeinated phase images. What is being compared is the feature-classifier’s performance when two different ground truths are used. For the first case, the ground truth results from the radiologist manually highlighting vessels in the uncaffeinated phase images. And for the second case, the ground truth results from the combined uncaffeinated and caffeinated highlighted images.
Figure 4.9 Color coded image super imposed on anatomical phase image of subject 1 slice 3 using uncaffeinated ground truth.

Figure 4.10 Color coded image super imposed on anatomical phase image corresponding to Figure 4.9 using combined ground truth.
Figure 4.11 Zoom-in version of color coded image super imposed on anatomical phase image of subject 1 (a) A zoomed-in version of Figure 4.9. (b) A zoomed-in version of Figure 4.10.
There are pixels colored-coded green in Figure 4.11a that are colored-coded yellow in Figure 4.11b. This suggests for the case of using the uncaffeinated ground truth images, the classifier detected the presence of vessels that were originally missed by the radiologist when highlighting the uncaffeinated images. But were latter detected by the same radiologist when highlighting the caffeinated image data, and showed up in the combined ground truth images.

Figure 4.11 is a typical illustration of the classifiers performance when uncaffeinated ground truth versus combined ground truth images for subject 1. For the case of subject 1, the overall performance of the classifier did decrease as indicated by Figure 4.8, but the classifier had some success in detecting previously missed underlying vessels. In the next section we continue to test the features and classifier using subject’s 2 and 3 image data.

### 4.5 Performance Evaluation of Features and Classifier via ROC Area Index and Color-Coded Images Using Testing Data

In this section we test the performance of the features and classifier trained on subject 1 uncaffeinated data using image data from subjects 2 and 3. In particular, for each of the test subjects, the input into the trained classifier consisted of all five zones from all five slices from the uncaffeinated images of subject 2 and 3. There are two sets of ground truth data, one being the uncaffeinated mark-up images and the other being the combined ground truth, similar to the presentation in Section 4.4. Using network settings from training on subject 1 uncaffeinated images, the classifier was first tested using
subject 2 data. We calculated the ROC curve and color-coded image for each of the
ground truth definitions. We repeated for subject 3.

Figures 4.12 through 4.14 illustrate the classifier’s ROC performance for subject
2. For Figure 4.12 the ground truth is the radiologist hand-marked uncaffeinated
images. For Figure 4.13 we used the combined ground truth images from the radiologist,
and Figure 4.14 provides their comparison. Figures 4.15 through 4.17 illustrate the
classifier performance via color-coded images of slice 3 of subject 2.

Figure 4.12 The ROC curve for the subject 2 uncaffeinated ground truth image data. The
ROC area index is 0.7673
Figure 4.13 The ROC curve for subject 2 combined ground truth image data. The ROC area index is 0.7506.

Figure 4.14 Uncaffeinated truth ground Vs. Combined ground truth of subject 2.
Figure 4.15 Color coded image of super imposed on anatomical phase image of subject 2 slice 3 using uncaffeinated ground truth.

Figure 4.16 Color coded image super imposed on anatomical phase image corresponding to Figure 4.15 using combined ground truth.
Figure 4.17 Zoom-in version of color coded image super imposed on anatomical phase image of subject 2 (a) A zoomed-in version of Figure 4.15. (b) A zoomed-in version of Figure 4.16.
Figures 4.18 through 4.20 illustrate the classifier’s performance for subject 3 via its ROC curves. For Figure 4.18 the ground truth is the radiologist markings from the uncaffeinated images. Figure 4.19 used the combined ground truth images, and Figure 4.20 provides their comparison. Figures 21 through 23 illustrate the classifier performance via color-coded images of slice 1 from subject 3. Figures 4.24 and 4.25 shows the comparison of the classifiers for subjects 1, 2 and 3 via one plot of superimposed ROC curves for uncaffeinated and combined ground truths respectively.

![ROC curve](image.png)

Figure 4.18 The ROC curve for the subject 3 uncaffeinated ground truth image data. The ROC area index is 0.8136.
Figure 4.19 The ROC curve for subject 3 combined ground truth image data. The ROC area index is 0.7929.

Figure 4.20 Uncaffeinated ground truth Vs. Combined ground truth of subject 3.
Figure 4.21 Color coded image of superimposed on anatomical phase image of subject 3 slice 1 using uncaffeinated ground truth.

Figure 4.22 Color coded image super imposed on anatomical phase image corresponding to Figure 4.27 using combined ground truth.
Figure 4.23 Zoom-in version of color coded image super imposed on anatomical phase image of subject 3 (a) A zoomed-in version of Figure 4.21. (b) A zoomed-in version of Figure 4.22.
Figure 4.24 Comparison of subjects 1, 2 and 3 using uncaffeinated ground truth.

Figure 4.25 Comparison of subjects 1, 2 and 3 using combined ground truth.
The central goal of selecting a set of features and designing a classifier is for their combination to suggest proper actions when presented with novel patterns, that is, blood vessel-tissue pixels not yet seen. This is the issue of generalization. A feature-classifier combination that exhibits good generalization suggests that a particular combination properly estimates the true underlying characteristics of all blood vessels and brain tissue that will have to be distinguished. Training on subject 1 and having repeatable results for subject data not used in training indicates the feature-classifier combination is a reasonably generalized one, as shown in Figures 4.24 and 4.25. More training samples would allow for a better estimate of the underlying characteristics, for instance the probability distributions of the two categories. Although our training and testing data are limited, we can be satisfied that these initial experiments suggest our feature-classifier combination may have promising performances to novel patterns. This implies the approach to obtaining the feature-classifier combination used in this dissertation may lend itself well to being used in a bigger study (e.g. using additional subjects and radiologist).

Furthermore, observing Figures 4.11, 4.17, and 4.23, some of the false positives using the uncaffeinated ground truth images became true positives when the combined ground truth images were used. This suggests it may be possible for a computer to discriminate hidden vessels not detectable by humans. Hence, underlying blood vessels within a given image may not have been readily apparent to the observer (i.e. radiologist); nevertheless, their presence was detected through their statistical relationship with their surrounding voxels. Additional research involving larger data sets are needed.
to further support such an implication. Nevertheless, these observations support the
approach of using features able to measure the texture-context information in phase
images as inputs into a multilayer perceptron based classifier to automatically identify
microvessels.

The objective of this research was to perform a preliminary study to show
feasibility, and therefore we did not claim a completed pattern recognition system or
toolkit to be used by radiologist. Given the difficulty of hand-marking the vessels in the
phase images, it is likely that the classifier is working about as well as a human. Further
studies using more subjects and more radiologist markings could help confirm this.
During the course of this research it became very apparent that such a tool would be a
significant benefit to radiologists wanting to identify microvessels in MRI image sets.
For example, controlling the mouse with the precision to highlight identifiable
microvessels on a computer monitor proved to be extremely difficult and very tedious for
the radiologist. Also as mentioned above, some subvoxel microvessels were not readily
identifiable within images, and remained unidentified/unhighlighted. Thus clinicians and
researchers would benefit from a CAD pattern recognition system that provides reliable
automatic identification of microvessels.

In the next chapter, we discuss our conclusions and outline the implications of
this research. We also discuss how this project may be extended to achieve better
performance and additional functionality.
Chapter 5

Conclusions and Future Work

5.1 Summary

In this dissertation research we have presented an initial approach to ascertain which anisotropic texture features can be used as classifier inputs for automatic detection of blood vessels that may not be specifically resolvable in the MRI image datasets. In particular, we provided 6 texture features, chose a feedforward neural network architecture for the classifier, and evaluated its performance.

In Chapter 2 we provided an overview of the basic principles behind magnetic resonance (MR) angiography and susceptibility weighted imaging. In particular, we provided a working definition of magnetic susceptibility and its ability to characterize magnetic materials. We reviewed traditional MR angiographic methods and their inadequacies for the purposes of this research project, and discussed the BMS effect. We presented the basic principles involved in MR imaging, and the advantages of phase
imaging as opposed to magnitude imaging. Finally, we presented the relationship between local variations in magnetic susceptibility and the local variations in the static magnetic field for cylindrical geometries, and an overview of the partial volume effects.

In Chapter 3 we focused on the various components of pattern recognition systems that were relevant to this dissertation research. In particular, we outlined the preprocessing steps involved in turning raw phase image data into phase images more amenable to feature extraction. Then we reviewed the mathematical principles behind the texture features that attempt to distinguish blood vessels from the surrounding tissues. Next we gave an overview of the histogram analysis methodology, the feature selection process, and the general network architecture for the classifier.

In Chapter 4 we presented the results from evaluating the features-classifier combination using three different subjects and two ground truth data sets. We trained the classifier using 3 zones from each of the 5 slices of subject 1 (for a total of 15 zones), and performed an initial performance evaluation using the remaining two zones from each of the 5 slices of the subject 1 (for a total of 10 zones). Then we performed a formal performance test of the trained classifier using all data from subjects 2 and 3 (25 zones each). For all tests we used two different ground truths: 1) the ground truth resulting from the radiologist manually highlighting blood vessels in the uncaffeinated images, and 2) the ground truth resulting from combining the manually vessel mark-up images from the uncaffeinated and the caffeinated.
5.2 Conclusions of Research

The following list summarizes the main implications of this research.

- Implication #1: The 6 texture features and the classifier exhibited encouraging generalization. This implies that particular combination reasonably estimated the true underlying characteristics of all blood vessels and brain tissue that will have to be distinguished for new patterns. This also implies the approach to obtaining the feature-classifier combination used in this dissertation may lend itself reasonably well to be used in a bigger study (e.g. using additional subjects and radiologist).

- Implication #2: The results revealed some of the false positives using the uncaffeinated ground truth data became true positives when the combined ground truth data was used. This suggests it may be possible for a computer to discriminate hidden vessels not detectable by humans. Hence, underlying blood vessels within a given image may not have been readily apparent to the observer (i.e. radiologist); nevertheless, their presence was detected through their statistical relationship with their surrounding voxels. Additional research involving larger data sets are needed to further support such an implication. Nevertheless, these observations support the approach of using features able to measure the texture-context information in phase images as inputs into a multilayer perceptron based classifier to automatically identify microvessels.
5.3 Future Work

One approach to extending this research is to incorporate more subjects and radiologists into the study. We can compare and combine the ground truth images between the various radiologists and/or the same radiologist. This would allow us to compare the classifier versus the human observer more thoroughly. For instance, we can measure inter-and-intra observer variations, and test if the classifier agrees with the radiologist(s) on the same level as the inter observable or intra observable agreement.

We can optimize the imaging acquisition techniques to make the voxels more isotropic, thus making through plane information more relevant. For example, this would help make the use of 3D co-occurrence texture analysis more useful. For this project when computing the co-occurrence matrices, we used only a distance of 1 pixel and the four principle directions, namely, horizontal, vertical, left-diagonal, and right-diagonal for 2D. This work can be extended by going to 3D, incorporating greater distances, and using interpolation to sample at more directions.

We can also explore the use of more complex neural network architectures. For instance, using multiple hidden layers or radial basis functions for neurons. All of the steps mentioned above could lead to even better performance of the feature-classifier combination by providing us a better estimate of the underlying characteristics of the human brain anatomy.
Bibliography


