A Fully Automated Breast Density Computation and Classification Algorithm

by

Steven M. McAvoy

B.Sc. Hons., The University of British Columbia, 2010

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

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Interdisciplinary Studies)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

July 2013

Abstract

Breast cancer is the most common cancer in Canadian women and early detection dramatically increases a woman's chance of survival. Until recently, womens' ages were considered the single most influential risk factor for developing breast cancer. Today, the density of fibroglandular tissue within the breast is considered just as important a risk factor as age. Because of this, accuracy and consistency while estimating tissue density is paramount. Currently, radiologists use the BI-RADS classification system to place mammographic images into one of four different categories. However, inter-observer variance has been shown to be as high as 30% and the methodology can be highly subjective. Many computer vision algorithms have been developed to automatically quantify breast density but only a few of these algorithms take advantage of the latest digital mammographic imaging technology. One algorithm, specifically designed to use digital mammography images, is explored in detail. Its ability to quantify and classify fibroglandular breast tissue is demonstrated and its accuracy is shown to be consistent with experienced radiologists. Finally, a modification to dramatically improve the running time is shown to have minimal effect on the overall accuracy of the algorithm.

Preface

The study in this thesis was conducted with the approval of the University of British Columbia - British Columbia Cancer Agency Research Ethics Board (UBC BCCA REB) under the certificate number H07-01485.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to extend my deepest appreciation to my supervisor, Dr. Patricia Lasserre, for the guidance, dedication, and patience during my time as a graduate student. Her commitment to her students and passion for teaching continues to inspire me and future students.

Additionally, I would like to extend my appreciation to my co-supervisor, Dr. Rasika Rajapakshe, for his expertise, vision, and the support offered to me while conducting my research.

Next, I would like to thank the BC Cancer Agency for the opportunity which allowed me to work with such talented and dedicated medical professionals.

Finally, I would like to thank the Computer Science faculty members at UBC Okanagan for their support and commitment in aiding their students to achieve success.

Dedication

This thesis is dedicated to my wife, Michelle, who selflessly sacrificed so much to allow me to pursue my dreams. She has been my strength and my inspiration.

Chapter 1

Introduction

Breast cancer in Canada is the most common cancer among women and survivability depends largely on early detection of the disease [36]. While many factors can affect a woman's predisposition to developing breast cancer, breast mammographic density, the bright area in a mammogram image, has been shown to be a reliable risk factor since its discovery in 1976 [35]. Within the last decade, breast density has shown to be at least as significant a risk factor as a woman's age [8] - the most significant risk factor known to date.

Given breast density's recently discovered role as a primary risk factor for women, its measurement and quantification will add additional benefit to existing breast cancer screening techniques, such as screening mammography. Currently, legislation with respect to informing women of the risk associated with breast density exists within three states of the United States of America [16]. National legislation within the United States of America is expected within the next two to three years. In Canada, similar national legislation has been passed within the House of Commons and is currently awaiting Senate approval [22].

To determine breast density, trained radiologists must examine mammo-

gram images and make a visual estimation. The practice is highly subjective and visual estimations add additional time to a radiologists already high workload. Within British Columbia, breast density is not examined as part of the province's mammography screening program partly due to the cost involved in the additional work radiologists would have to perform. Given the upcoming legislation and imminent patient requests for breast density risk evaluation, a method of reliably automating the quantification of breast density using computer vision techniques is highly desired. Currently, many such computer vision algorithms exist with varying levels of reliability and accuracy, however most of these algorithms do not take advantage of new state-of-the art digital mammography devices which generate images with improved clarity and resolution. Of the few algorithms designed to use digital mammogram images, almost all require the use of the raw sensor data which results in an increase in digital storage costs if implemented.

This thesis investigates one such computer vision algorithm specifically designed to quantify breast density from digital mammogram devices without dependence of the raw sensor data. The image examined by a radiologist is the same image examined by the algorithm. To provide additional benefit from the original implementation, image classification size and its effect on the algorithms running time and accuracy are explored.

Chapter 2

Background

2.1 Digital Mammography

Mammography has been a medical imaging diagnostic tool since the late 1960s and has been used for cancer screening of the breast since at least 1976 [19]. Mammography is the process of imaging the internal structure of the human breast using low energy X-ray radiation. Until recently, all mammography devices recorded the attenuation of X-ray radiation from its source in an analog manner by using film dedicated for mammography. Modern mammography devices have replaced the analog film with digital sensors which record X-ray attenuation using discrete numeric values (i.e. digital).

X-ray radiation is harmful to biological tissue and therefore mammography devices have always aimed to deliver the minimal X-ray dose necessary for the image clarity required for diagnosis. Full-field digital mammography ($FFDM$) can achieve higher image resolution and clarity using X-ray doses which are approximately 40% lower than their analog counterparts. Typical analog mammography devices deliver an average dose of approximately 2.0mGy to the breast while digital mammography can achieve higher quality imaging using a substantially smaller dose of around 1.2mGy.

The human female breast is comprised of many types of tissue as shown in Figure 2.1. Tissue types include the chest wall, pectoralis muscles, lobules, nipple, areola, milk ducts, fatty tissue, and the surrounding skin. Throughout the breast, blood vessels, veins, and connective ligaments are intermingled with the surrounding internal tissue. For the purposes of breast density, only the lobules, milk ducts, and connective ligaments are considered and are collectively referred to as *fibroglandular* tissue.

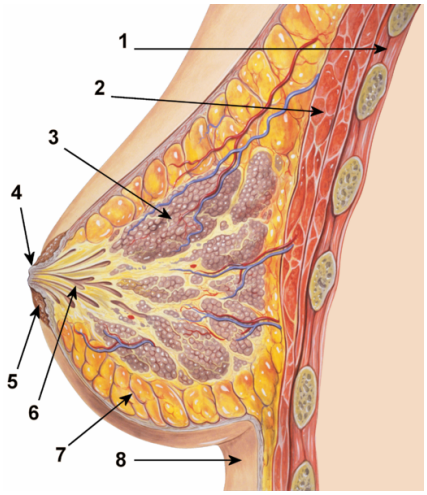


Figure 2.1: Anatomy of the human female breast [20].
1) Chest wall, 2) Pectoralis muscles, 3) Lobules, 4) Nipple, 5) Areola, 6) Milk duct, 7) Fatty tissue, 8) Skin.

Since fibroglandular tissue is more dense than the surrounding fatty tissue, it attenuates X-rays at a higher rate and appears as regions of gray and white on both analog and digital mammogram images as shown in Figure 2.2. Conversely, fatty tissue appears as dark.

During a typical mammographic exam, the breast is imaged using two different angles of view as shown in Figure 2.3. The first angle, cranial-caudal

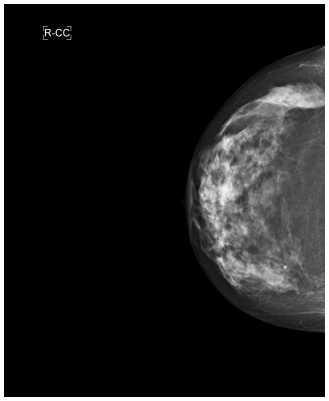
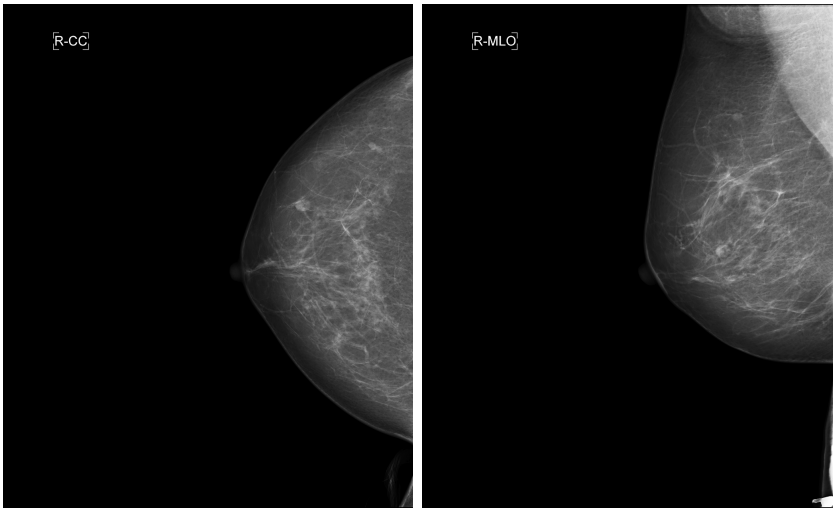


Figure 2.2: A typical digital mammogram image.

or abbreviated *CC*, is obtained through an axis which exists between the patient's head and feet. The second viewing angle is offset at approximately 45° to the cranial-caudal axis. These two viewing angles assist radiologists in determining the type and orientation of the internal breast tissue.

Unlike analog mammography devices which required medical professionals to view the film on a bright view box, FFDM devices display their output on a wide variety of high-resolution computer screens. Since the digital mammography image capture process is entirely digital, device vendors have incorporated various image processing techniques to optimize the image display for examination by the human eye. To accomplish image enhancement, FFDM devices perform two operations in sequence: 1) acquisition of the X-ray attenuation values via the digital sensors which results in the raw pixel data corrected for sensor response and noise, and 2) the application of various proprietary image enhancement algorithms on the raw data which results in the final enhanced image. The raw pixel data image is never viewed directly by medical professionals and exists only temporarily as a



(a) Cranial-caudal (CC) (b) Mediolateral-oblique (MLO)

Figure 2.3: Mammogram image orientations.

means to create the final output image - the processed or diagnostic image.

The application of image enhancement algorithms such as histogram equalization or image sharpening always results in a loss of original sensor data. Generally, performing quantitative image analysis of any kind, including a breast density computation algorithm, would utilize the raw pixel data image. However, since the raw pixel data is only an intermediate step in the FFDM output process and a diagnosis is never extracted directly from it, these images are not stored for any significant length of time and are never linked to a patient's electronic medical record.

2.2 Breast Density

Breast density is defined as the ratio of the amount of fibroglandular tissue within the breast to the amount of total breast tissue. This ratio is typically represented as a percentage using Equation (2.1).

$$\%_{density} = \frac{Area_{fibroglandular}}{Area_{breast}} \times 100 \tag{2.1}$$

Today, breast density is merely estimated by trained radiologists who typically spend under 90 seconds reviewing a single mammogram image within a viewing environment. The results are widely subjective and can vary greatly between other radiologists by as much as 30% [14]. This large inter-observer error indicates radiologists lack a common definition of dense tissue. Additionally, intra-observer error can also differ greatly by as much as 20% suggesting that even experienced radiologists have difficulty maintaining a consistent opinion of what constitutes dense tissue. Moreover, no standard methodology or technique has been officially adopted by the medical community making density value comparisons between radiologists even more difficult to interpret.

To aid in reducing the subjectivity and error, the Breast Imaging-Reporting and Data System, or *BI-RADS*, was introduced in 1998 [32]. The BI-RADS tool presents four distinct categories for breast density as shown in table 2.1.

Instead of quantifying breast density by estimation, radiologists select the category which best describes the mammogram image being examined. While effective, the BI-RADS scale has not completely eliminated subjectivity, especially between categories II and III where disagreement between

Table 2.1: The BI-RADS breast density categories.

BI-RADS Category	*Description*
I	Almost no fibroglandular tissue
II	Scattered fibroglandular densities
III	Heterogeneously dense
IV	Extremely dense

radiologists can be as much as 35% [2, 7, 9, 27]. Repeat readings by the same radiologist at different times has shown to be more consistent, however, the intraobserver error can be as high as 10% [14].

Within the medical industry, algorithms which produce reliable and deterministic results are preferred so as to limit the exposure of legal liability. For the purposes of breast density computation, a deterministic algorithm is a primary consideration to ensure that historical values can be reproduced in the future in the event of any litigation or investigation.

Many approaches to automated computation have opted to use the raw pixel data as the primary data source [1, 11, 28, 30, 33] due to its characteristics of possessing the most information possible for analysis. Unfortunately, the raw pixel images are almost never stored with a patient's electronic medical record giving them limited capability in assessing breast density retroactively.

Since the diagnostic image from FFDM devices is the only image stored long-term for legal reasons, breast cancer risk for women who have received digital mammograms could be ascertained by developing a software algorithm which could compute breast density from the diagnostic images instead of the raw pixel images.

2.3 Previous Attempts at Automation

Over the last 20 years, many attempts have been made to automatically quantify and classify breast density from mammogram images. However, since digital mammogram devices have only been a widely accepted modality within the last 5-7 years, most image analysis and density quantification algorithms were designed and developed for use on digitized mammographic film.

Since initial research began on automated breast density, digitization of analog sources has improved significantly both in image resolution and digitalization techniques. This evolution is reflected in the methodologies explored to determine density.

Three fundamental image processing techniques have been the primary focus for advancing accuracy in breast tissue density quantification: statistical analysis, the use of calibration between pixel gray-scale values and X-ray attenuation, and finally segmentation via pixel clustering.

2.3.1 Image Thresholding - Statistical Analysis

Perhaps the most widely-used and oldest image processing technique is the use of statistical analysis of the entire image or its components. Since any digital image is essentially a 2-dimensional array of values, many statistical techniques can be employed to gain insight into the contents of the image. Of significant importance in statistical analysis, variance is commonly utilized in the examination of pixel intensity values. Early exploration into automated breast density quantification examined the variance of pixel values

within many small regions of the breast area and compared it to the global variance of the breast itself [13]. Regions with a variance which exceeded the global average were deemed dense tissue and regions with variances below the global average were labelled as fatty. This method successfully quantified breast density to within accepted inter-observer error 85% of the time. Another approach [31] applied Kittler's minimum error thresholding algorithm [15] to digitized mammographic films and yielded similar results.

The analysis of an image's histogram has also shown to reveal possible insight into breast density [6, 18, 29, 36]. Fatty tissue tends to give image histograms a leftward skew while dense tissue skews the histogram to the right. Using this information, the image can be analyzed at varying levels of resolution to determine which components are dense or fat.

One of the most popular techniques employed by researchers when striving to quantify breast density is to attempt to calibrate the pixel intensity value to the amount of X-ray attenuation. Achieving this objective involves imaging a plastic device, known as a step wedge, which contains varying levels of material density. An X-ray image is then acquired via the mammography machine and its resulting film is then digitized using a film scanner. Once the image has been digitized, pixel intensity values on objects within the image can be mapped to the known attenuations, thus achieving a rough calibration between matter density and pixel intensity. This process is time consuming and must be performed manually on each mammogram machine, even on machines which share the same manufacturer and model number due to the small fluctuations in delivering an X-ray dose.

2.3.2 X-Ray Calibration

The use of calibration alone is insufficient for the determination of breast density. Additional supporting techniques are required as calibration does not determine at which point tissue stops being fatty and starts becoming fibroglandular. The difficulty in determining the density threshold is compounded when the differences in anatomy between women coupled with the differences in X-ray energy used for different breast sizes is considered.

One of the earliest attempts at density quantification used pixel intensity calibration along with several other statistical properties of the mammogram image [4]. The authors compared supporting methods to pixel intensity calibration by examining statistical characteristics such as gray-level average, standard deviation, breast width, breast height, and coefficient of variance among others. In all, approximately 200 statistical features where used and compared. While no single feature could fully correlate with breast density, multiple features were combined to enhance the accuracy so that 88% of the images examined would fall within the expected inter-observer error for breast density value.

More recently, research into refining the methods and processes of pixel calibration has yielded small improvements in accuracy [11, 12]. Heine has shown that removing the outermost 25% of the breast tissue before calibration is applied increases accuracy slightly by ensuring that only the areas of the breast which come into complete contact with the compression paddles are used. Areas of the breast which are not under compression contain X-ray attenuation induced by the surrounding atmosphere and thus introduce

additional error into the calibration. Heine's work has also been applied to modern FFDM images, however, his work was restricted to the sole use of raw pixel data.

While image analysis techniques like thresholding, clustering, and X-ray calibration have all yielded excellent results when used on digitized mammographic films, each technique is rendered unsuitable for use when analyzing post-processed FFDM images. Post-processed FFDM images are optimized for the human eye and therefore have had one or more proprietary image processing techniques applied to them.

One common processing technique used by manufacturers of mammogram imaging devices is histogram equalization which distributes a small range of gray-level values across the entire range of available gray-level values. This distribution results in increased contrast allowing the human eye to respond to subtle changes in gray-level values. In addition, manufacturers will further alter the contrast of FFDM images by tailoring each mammogram imaging device's output image to the preference of the reading radiologist who interprets the images. This could lead to increased inter-observer error in cases where two or more radiologists are determining breast density from the same digital mammography device.

The combination of these two techniques makes the use of thresholding (either by statistical analysis or via clustering) extremely difficult as the contrast of dense tissue will differ, not only from machine-to-machine, but also image-to-image. Additionally, an unsupervised thresholding algorithm has much less pixel value range available in which to derive an appropriate threshold value than if it were using the unprocessed, or raw, image data.

Using an X-ray calibration technique is theoretically possible with FFDM, however, given that each imaging machine produces a unique output due to the operator's contrast preferences, each machine installation would require its own calibration. Moreover, each time the machine's output contrast was altered, another calibration would be required. This becomes laborious and is further complicated when taking into account the legal requirements for the medical industry. If breast density algorithms based upon X-ray calibration are to be consistent with previous results, then each machine must not only store its current calibration data, but the calibration data for its entire useful lifespan.

2.3.3 Image Thresholding - Clustering

Another image segmentation technique, known as clustering [17], has also been explored for the use in determining mammographic density. With clustering, pixels are grouped into a pre-set number of clusters. For the purposes of determining breast density, 2 clusters are typically employed representing fatty and fibroglandular tissue, respectively. K-means clustering has proven to be the most popular statistical algorithm for successfully clustering data and is generally used in most image processing applications.

To compute which pixels belong to which cluster, the k-means algorithm starts by selecting 2 random values and pixels are grouped into one of the two clusters by determining which of the two clusters gives rise to the smallest difference. Once all the pixels have been clustered, the algorithm computes the centroid value of each cluster and again, the pixels are assigned to the cluster giving the smallest difference. The algorithm repeats

until cluster membership no longer changes. While effective at clustering pixels containing similar intensity values, the k-means algorithm falls short when attempting to group data with more complex attributes, such as distinguishing between veins and fibroglandular tissue.

Research into the use of clustering for determining mammographic density has shown that simply applying the k-means algorithm is insufficient [23, 24]. The optimal results were obtained when using additional statistical characteristics of the image and by modifying the k-means clustering algorithm to use cluster membership probabilities. This altered algorithm is commonly known as c-means clustering [3]. Using the combined c-means and statistical properties, breast density has been successfully computed to within iter-observer error 86% of the time [3, 23, 24].

While each of these image processing techniques have achieved a wide level of adoption among image processing applications, they all fall short of one key requirement for segmenting FFDM images for the purposes of computing breast density: classification of fibroglandular tissue. Since not all tissue within the breast is considered dense tissue, a method of classifying and identifying the correct tissue type is needed.

Historically, thresholding image segmentation techniques have been the primary tool for developing an automatic breast density algorithm. In 2010, a breast tissue segmentation technique which did not rely on thresholding was developed by Oliver et al. [26]. This new technique adapted a widely-used facial recognition algorithm, known as Eigenfaces [34], to identify dense and fatty tissue. In addition, Oliver's algorithm was able to determine and omit areas of the breast which were not classified as fibroglandular

tissue while analyzing only the diagnostic FFDM image data. In this thesis, Oliver's algorithm is explored and its sensitivity to various parameters are examined in great detail.

Chapter 3

Automatic Breast Density Algorithm

Analyzing any digital mammogram image for breast density is essentially a three step process:

1. Isolate the breast tissue from the image background;
2. Segment the breast into either dense and fatty tissue;
3. Compute the ratio of dense tissue to total breast tissue.

The first step in the algorithm can be achieved using basic image processing techniques such as histogram analysis. However, since the mammogram images being analyzed in this thesis have undergone image enhancement processing, the background has already been removed leaving only the breast tissue behind.

The third step is trivial once steps 1 and 2 have been completed. The dense pixels are counted and divided into the total number of pixels making up the entire breast.

For the second step, a supervised algorithm, based on Oliver's modified Eigenfaces implementation, was chosen to determine which tissue pixels were

dense and which were not.

3.1 Algorithm Overview

Figure 3.1 illustrates the general idea of the algorithm in terms of its inputs and outputs as defined by Oliver. Once trained by the training images, a standard FFDM is analyzed as input by the algorithm. The resulting output is a binary image containing pixel values of only zero for non-fibroglandular tissue and one for fibroglandular tissue.

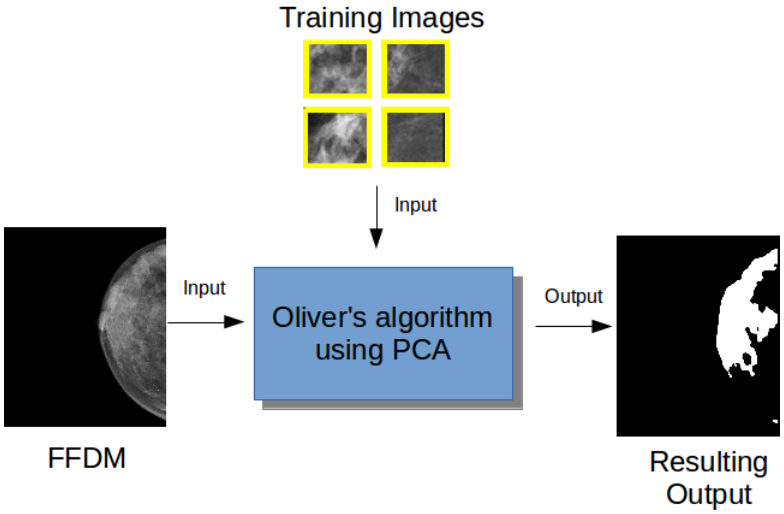


Figure 3.1: Inputs and outputs of Oliver's algorithm.

The training images are created by dividing the pixels which make up the breast tissue on each mammogram image into 50×50 pixel patches. For each patch, a trained radiologist assesses whether or not the patch is classified as either dense or fatty tissue. Each patch is then transformed into a lower dimension subspace using *Principal Component Analysis* (PCA). This

process is repeated for numerous mammographic images until a sufficient definition of dense tissue has been developed.

To assess breast density, the mammogram image is analyzed at each pixel location which represents breast tissue. At each pixel location, the surrounding 50×50 region is extracted and compared against the training images using PCA and the determination of either dense or fatty tissue is made via K-nearest-neighbour, as shown in Figure 3.2. After each pixel location has been analyzed, a binary image containing a map of dense and fatty tissue is created. The resulting density is then computed by summing the number of pixels which make up the dense tissue and dividing it into the total number of pixels which represent the entire breast area.

Figure 3.2 describes the core algorithm as defined by Oliver. Since the algorithm being used is supervised, it must be given a set of criteria which define what is meant by dense breast tissue. To accomplish this, a series of training images are required along with their respective classification as either dense or fatty.

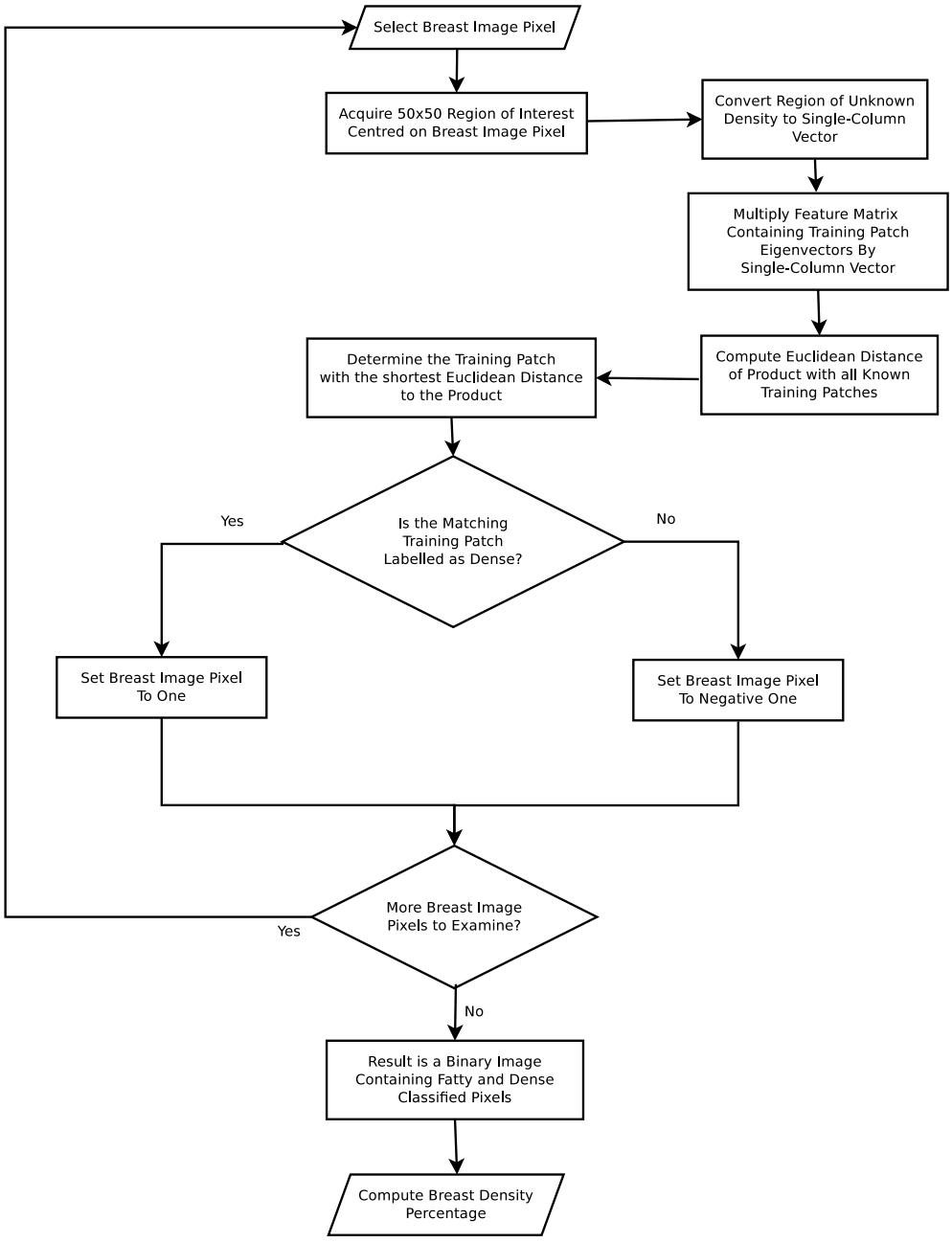


Figure 3.2: Core algorithm for computation of breast density.

3.2 Principal Component Analysis

Principal Component Analysis, or *PCA* as it is commonly referred to, is a statistical technique used to find patterns in data of high dimensionality. Seemingly uncorrelated data is transformed into a subspace where possible correlations can be determined by analyzing which components have the strongest influence on statistical variance. The variable which causes the greatest variance is known as the principal component.

Analysis using PCA is performed on single dimensional vectors. While digital images are typically represented as two-dimensional matrices for mathematical convenience, the representation of the image data in a computer's memory is always a single-dimension vector. To provide a consistent transformation, Equation (3.1) is used to convert an image's Cartesian coordinates to a single-dimensional vector where I is the matrix of pixel intensity values, x and y represent the pixel coordinates where the origin begins at the upper left-most corner and their values increase to the right and downward respectively, and i is the pixel intensity value at the specified coordinate.

$$I(x,y) = \begin{bmatrix} i_{1,1} & \cdots & i_{1,x} \\ i_{2,1} & \cdots & i_{2,x} \\ \vdots & \ddots & \vdots \\ i_{y,1} & \cdots & i_{y,x} \end{bmatrix} = \vec{p}(n) = \begin{bmatrix} i_{1,1} \\ \vdots \\ i_{1,x} \\ i_{2,1} \\ \vdots \\ i_{2,x} \\ \vdots \\ i_{y,x} \end{bmatrix} \tag{3.1}$$

In the context of this breast density quantification algorithm, an image patch is a 50×50 pixel region from the breast area. To build a training set for which future patches of unknown classification can be compared against, patches with known density are first converted into their respective single-dimensional vectors using Equation (3.1). After each image has been converted, a *mean* patch image is produced so that differences in patches can be assessed.

The mean patch image is found using Equation (3.2), where $\vec{p}$ is a single-dimensional vector comprised of a patch's pixel intensity values and M is the number of patches with a known density classification which make up the training set.

$$\vec{\mu} = \frac{1}{M} \sum_{k=1}^{M} \vec{p_k} \tag{3.2}$$

With the average patch defined, a covariance matrix can be constructed using Equation (3.3). The covariance matrix allows for comparisons of one

patch against all other patches by generalizing the notion of variance to multiple dimensions.

$$C_v = \sum_{k=1}^{M} (\vec{p_k} - \vec{\mu})(\vec{p_k} - \vec{\mu})^T \tag{3.3}$$

Since the covariance matrix C_v, is square, the unit eigenvectors and corresponding eigenvalues can be determined. Each unit eigenvector is orthogonal and highlights related data in each dimension. The corresponding eigenvalues indicate the strength of the relationship with the highest eigenvalue associated with the principal component of the data set. Not all possible eigenvectors are necessary to record the most significant components. Using the most significant eigenvectors, as indicated by their eigenvalues, a feature matrix W can be obtained as shown in Equation (3.4). P eigenvectors are possible from a total of N training patches, where $P \leq N$. To achieve optimal classification, the exact value of P must be determined empirically by ranking the eigenvectors by their eigenvalues and selecting an optimal portion of those containing the highest values. With P determined, the feature matrix, W, can be used to classify patches of unknown density by transforming these patches into the PCA multi-dimensional space via Equation (3.5), where p is a patch of unknown density having identical dimension to a training patch.

$$W = \begin{bmatrix} \vec{w_1} & \vec{w_2} & \cdots & \vec{w_P} \end{bmatrix} \tag{3.4}$$

$$\vec{y} = W^T \vec{p} \tag{3.5}$$

The training patch with the closest Euclidean distance is chosen as the correct classification using Equation (3.6), where $\vec{y_t}$ and $\vec{y_u}$ are the single-column eigenvector representing respectively a training patch and a patch of unknown density which were transformed via Equation (3.5). Once the training patch with the closest Euclidean distance is found, its classification is given to the individual pixel located at the center of the patch of unknown density.

$$d_i = \min_{i=1}^{N} \sqrt{\sum_{k=1}^{M} (y_{t_{i_k}} - y_{u_k})^2} \tag{3.6}$$

As an example, a patch of unknown tissue density having a size of 50×50 pixels is converted into a single-column vector, $\vec{p}$, by means of Equation (3.1) having a resulting dimension of 1×2500. Once converted, the vector $\vec{p}$ is then transformed into the PCA multi-dimensional space via Equation (3.5) resulting in $\vec{y_u}$. Using Equation (3.6), the smallest Euclidean distance between all training images, y_t, and the patch of unknown density, y_u, can be determined. The training image which produced the smallest distance is the one that is most similar in features to the patch of unknown density. The patch of unknown density is then classified with the same classification as the selected training image.

3.3 Computation of Breast Density

With the ability to classify pixels as either dense or fatty, the entire mammographic image can now be analyzed. At each pixel, a classification

is determined and the corresponding pixel is set to a value of 1 if it is dense or 0 if it deemed fatty. The end result is a binary image comprised of pixels having only the value 1 or 0. The density is computed as a percentage by summing all the pixels having a value of 1 and dividing it by the total number of non-zero pixels in the binary image.

While Oliver does not state the running time for his algorithm, it is evident from the repeated PCA transformations which occur on a pixel-by-pixel basis that the computational expense is costly.

Chapter 4

Materials and Methodology

4.1 Ground Truth Observations

To provide a baseline for which all computed breast density values could be compared against, observations from trained radiologists is required. Providing these observations were Dr. Paula Gordon from the British Columbia Women's Hospital, Dr. Stuart Silver from the Victoria General Hospital, Dr. Catherine Staples from the Kelowna General Hospital, and Dr. Stacey Piche from the Penticton Region Hospital. Each physician was a trained radiologist with numerous years of experience examining mammogram images and estimating breast density.

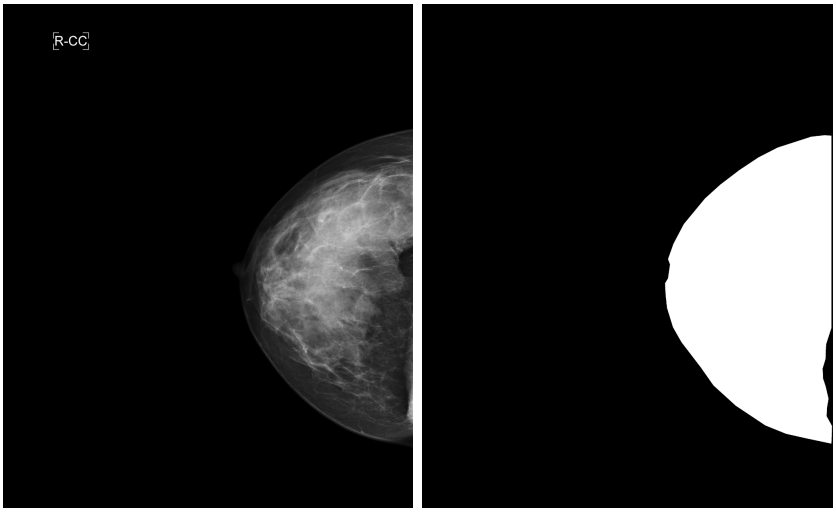
Observations from each radiologists were recorded using the Cumulus 4.0 [5] user-assisted thresholding software. Cumulus 4.0 requires its users to manually input the required threshold value for dense tissue by means of manipulating a slider tool which highlights the dense tissue. This allows the users to continually adjust the input value until the most appropriate threshold value is obtained. Once this value is found, the software then stores the threshold value in an internal database and continues to the next image. In all, each radiologist individually evaluated 66 FFDM images and their corresponding threshold values and tissue densities were recorded.

To establish a ground truth baseline for which the automated density algorithm could be compared against, the average of each of the 4 independent observations from each FFDM was computed. The independent observations for each radiologist were comprised of the average from one set of FFDM images presented to the radiologists twice in a randomly generated order to account for intra-observer error and any possible bias in the image order.

4.2 Training Database

The core of the automated density algorithm relies on the Eigenfaces algorithm which, in turn, relies exclusively on PCA. In order for the Eigenfaces algorithm to determine which parts of an image are dense and which are fatty, a definition for dense tissue was needed to which unknown tissue could be compared. This was accomplished by creating a set of training images which defined both dense and fatty breast tissue.

In Oliver's original implementation, training images were selected in a manual fashion by trained radiologists. However, to incorporate a more consistent selection criteria, the training patches for this experiment were selected using the recorded data stored in Cumulus during each of the participating radiologists' observations. In total, 66 FFDM images were successfully observed by all radiologists and 6 images were selected as suitable candidates from which training images could be harvested. Of the chosen 6 FFDM images used for training, 2 were considered typical fatty breasts, 2 were considered medium density breasts, and the remaining 2 were consid-



(a) Cranial-caudal (CC) (b) Associated Image Mask

Figure 4.1: Mammogram image orientations.

ered typical high density breasts. All 6 images were chosen after consultation with each of the radiologists and were free from anomalous artifacts such calcifications, atypical tissue, and lesions.

Prior to selecting the training images from the candidate FFDM images, pixels representing unsuitable or non-fibroglandular tissue were required to be removed. Such areas include the nipple and surrounding region, as well the area immediately inside the skin-line boundary. In the original implementation, Oliver uses a previously developed automated image processing technique [25] to complete this step. However, for the purposes of this experiment, the non-breast tissue removal process has been replaced by the application of a manually procured image mask, as shown in Figure 4.1, which aids the algorithm in identifying only breast tissue.

Training images were created from the candidate 3328×4084 pixel FFDM images by dividing the pixels containing breast and fibroglandular tissue into many 50×50 regions. Each region was then segmented by extracting the threshold pixel value stored in the Cumulus database recorded by each each radiologist for each corresponding FFDM image. The ground truth pixel intensity threshold value was computed by taking the average recorded threshold values from each radiologist's observation. If the number of pixels above the threshold value within the region was greater than 50% of the total number of region pixels, then the region was classified as dense. Conversely, regions which had less than 50% of their pixels above the threshold value were classified as fatty.

Once training images had been extracted from all 6 candidate FFDM images, it was necessary to reduce the training set down from 50,000 to a smaller value to reduce the overall processing time for each image. Oliver empirically determined an optimal number of 108 when implementing his algorithm. From the 50,000 training patches, 3 groups were created: 1) a random sample over the entire training image population; 2) a random sample of training images which were either extremely dense or extremely fatty; 3) a random sample of training images which were either slightly dense or slightly fatty. For each of the 3 groups 60 dense and 60 fatty training images were chosen.

4.3 Computation of Breast Density

Computation of the breast density on a mammogram image with unknown density begins by examining each pixel of the image. As each pixel is visited, a surround sample region centred on the pixel is extracted having the same dimensions as the training patches (50×50). This region is then transformed into a multi-dimensional subspace via PCA and is compared against the set of training criteria. The determination of each pixel being dense or fatty is made by measuring the Euclidean distance of its corresponding region to each training patch within the subspace and selecting the classification of the training patch which has the shortest distance. This process is then repeated for each pixel of breast tissue within the mammogram image.

After each pixel has been analyzed, a new binary image is created containing values of 0 for fatty tissue and values of 1 for dense tissue. Computing the density percentage is achieved by counting the number of dense pixels and dividing it by the total number of both fatty and dense pixels as shown in Equation (2.1).

To investigate the algorithm's accuracy performance, breast density was also calculated by varying the classification size for which tissue was classified. For instance, density classification set larger regions of sizes 3×3, 5×5, 10×10, 20×20, , 30×30, 40×40, and 50×50 instead of the original implementation which classified pixel-by-pixel. Throughout these experiments, the training patch image sizes remained constant.

Chapter 5

Discussion of Results

A breast density algorithm's success can be measured in many different ways. Of primary interest to this study, was the algorithm's performance compared to that of well-trained and experienced radiologists as well as the effect of reducing the running time on accuracy.

5.1 Training Images

The density algorithm requires training images to be loaded prior to performing density computations on mammogram images of unknown density. The number of training images used can affect the accuracy of the algorithm significantly. Too few training images result in a poor definition between fatty and dense tissue resulting in incorrect classifications. The number of training images required to ensure optimal algorithm accuracy is remarkably low. Oliver described the use of 108 training images to achieve optimal performance. The data in this experiment found an optimal number of 120 images as shown in Figure 5.1. While the value of 120 was the empirically determined optimal value for this experiment, the net benefit in algorithm accuracy performance between 72 and 120 training images decreases the average absolute error by only 0.02%. Additionally, using more

than the optimal number of training images did not yield any measurable increase in accuracy performance.

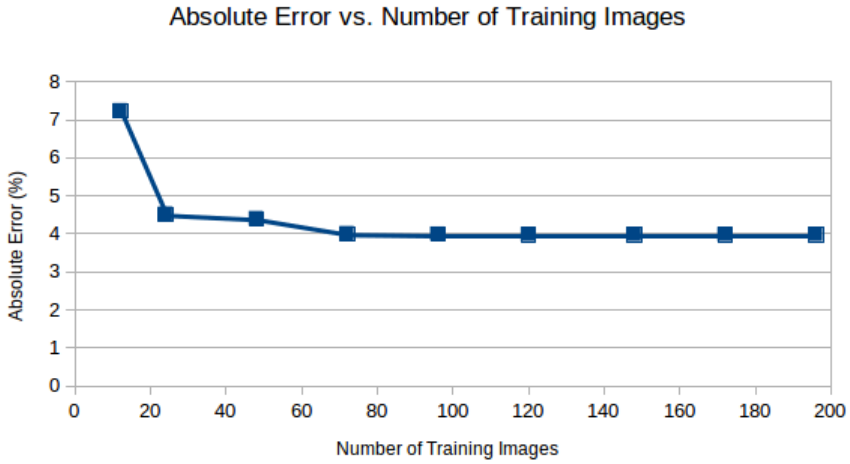


Figure 5.1: Effect of training image quantity.

Table 5.1: Effect of patch selection on algorithm accuracy.

Selection Type	*Ground Truth Correlation*
Random Selection	0.871
Extreme Densities	0.904
Similar Densities	0.903

The type of patch images selected for training did produce a small influence on the algorithm's overall accuracy. Table 5.1 displays the results of 3 sets of 120 training images. The first set was comprised of a random sample of 10 dense regions and 10 fatty regions for each of the 6 training mammograms. The resulting 120 images represented a random distribution of varying levels of fibroglandular density for both dense and fatty tissues. The second set selected training patches from regions where the dense and

fatty tissue was well-defined and their corresponding locations were in areas of maximal or minimal density, respectively. Finally, the third set selected training patches from regions where dense and fatty tissue contained gray-scale pixel values which were similar in value.

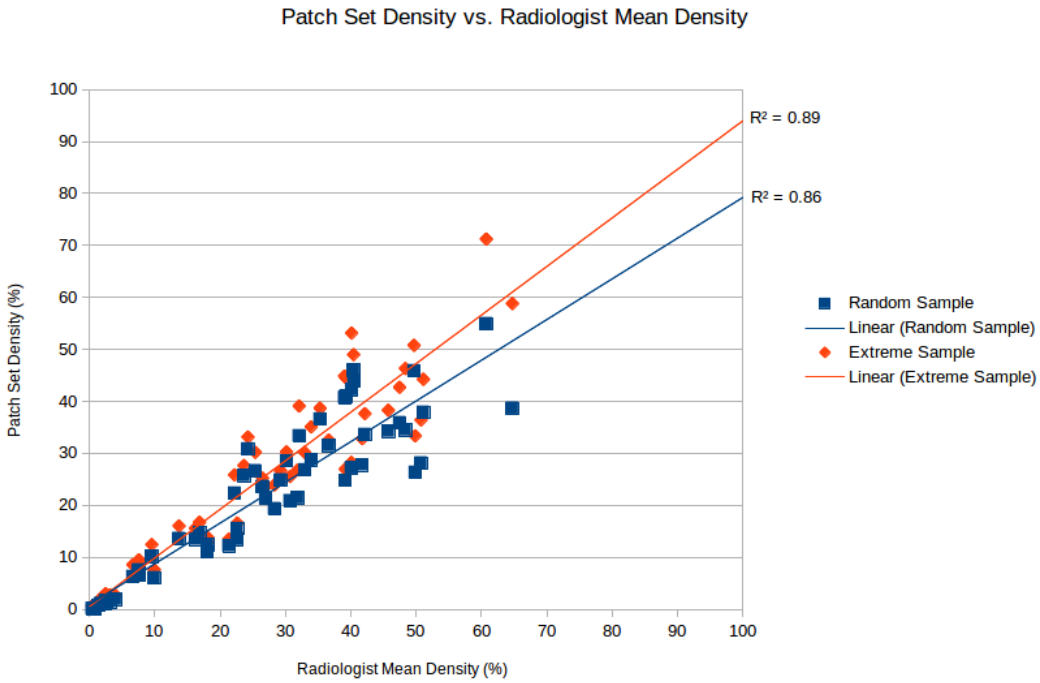


Figure 5.2: Training patch set accuracy.

Overall, the accuracy of the algorithm's performance was affected minimally by the types of patches used for training. However, the optimal configurations were achieved when using images which represented either the areas of greatest density extremes, or areas of density ambiguity. Due to the fact that the training set with randomly sampled densities did not depict fibroglandular tissue as accurately as a set containing densities at

their respective extremes, the resulted computed density values were lower than the observed densities along with a slight increase in variability from observed values, as shown in Figure 5.2.

5.2 Ground Truth Density

To establish a baseline for breast density comparison, four participating radiologists were asked to observe and compute the breast density on a set of FFDM images. Each radiologist viewed the set of FFDM images twice in a randomly generated order to minimize the effects of intra-observer error and possible image order bias. The average of the two observations was then used as a ground truth observation from the respective radiologist.

As shown in Figure 5.3, the radiologists' observations were in general agreement for the large majority of observed images. Overall, the radiologists' average inter-observer absolute error was 11.5% which falls within the expected range found within relevant literature. Figure 5.4 displays the average absolute error plotted against breast density which clearly shows that FFDM images containing high levels of fibroglandular tissue density induce an increase in inter-observer error.

Figure 5.5 shows the average breast density computed by all four radiologists' observations for each FFDM image observed. These values formed the ground truth density for each FFDM image from which the algorithm's performance could then be compared against.

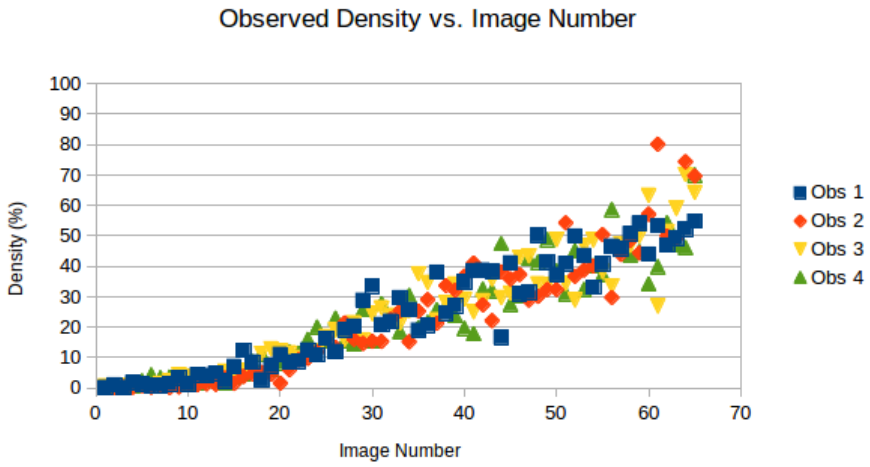


Figure 5.3: Radiologist observations.

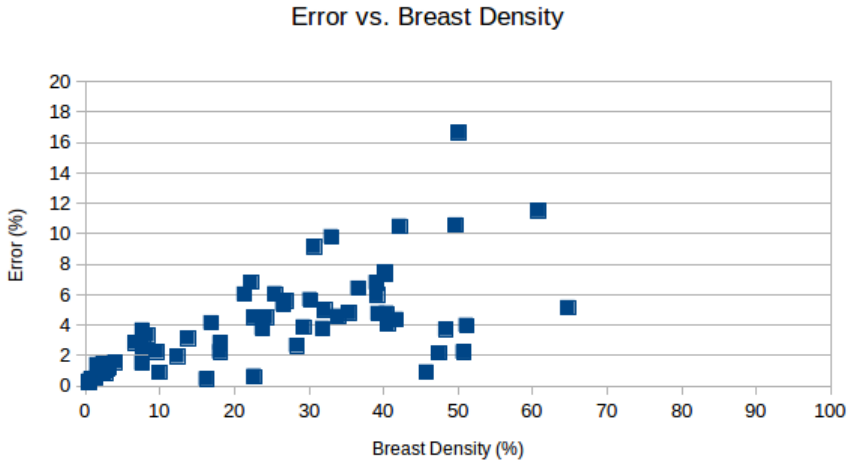


Figure 5.4: Radiologist average error.

5.3 Number of Principal Components

Another key component of the algorithm is the number of principal components used when performing the Principal Component Analysis to com-

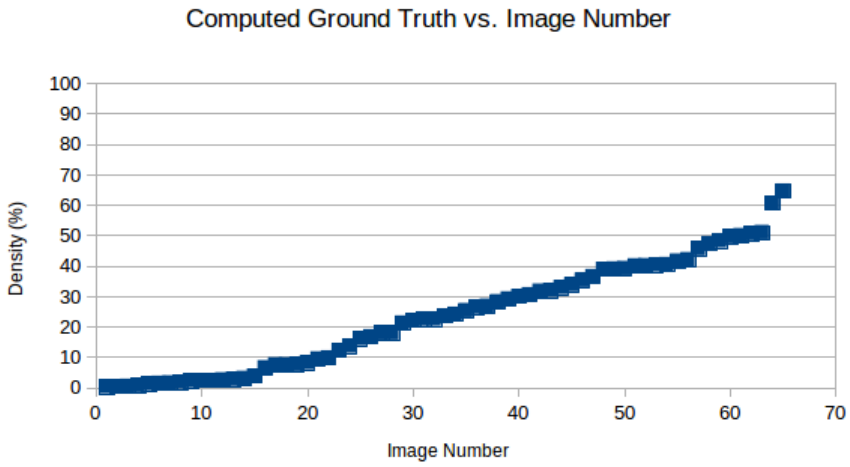


Figure 5.5: Radiologist ground truth.

pare patches of unknown fibroglandular density with the training patches. As mentioned in Section 3.2, the number of principal components, P, must be determined empirically. Due to the fact that the algorithm's author does not reveal the number of principal components used to generate his results, the optimal number for P was determined experimentally for the generation of all results in this thesis.

Figures 5.6 and 5.7 show the relationship between the number of principal components chosen for the Eigenfaces algorithm and its effect on the correlation between computed density and the ground truth density. The optimal number of principal components was found to be 10. Increasing the number of principal components beyond this value produced an extremely minute, but measurable, decrease in accuracy. However, even when using more than 100 principal components, the accuracy never dipped below 0.3% of the optimal value.

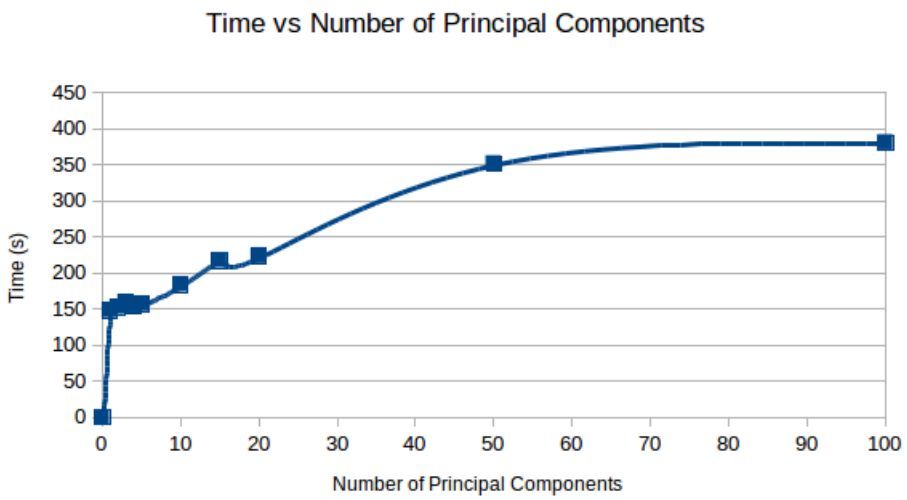


Figure 5.6: Effect of number of principal components on time.

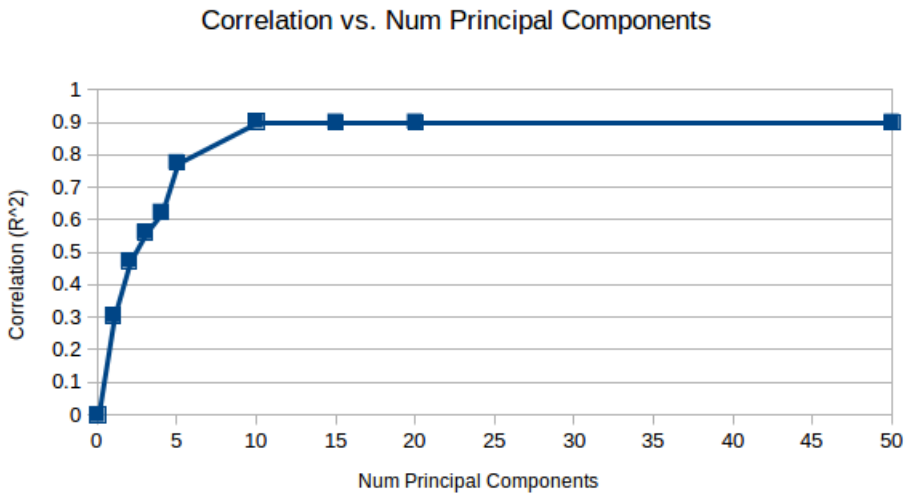


Figure 5.7: Effect of number of principal components on accuracy.

Optimizing the processing time for the density computation algorithm was a secondary area of interest. Since the original author did not provide performance metrics, determining the optimal configuration for both accuracy and performance was explored. Predictably, processing time for the algorithm was increased when the number of principal components was raised. Figure 5.6 shows that after approximately 45 principal components are used, the time increase begins to plateau. The reason for such a plateau are currently unclear but may have to do with the third party library's PCA implementation since the maximum number of Eigenvalues, or principal components, possible is much larger than the range examined. Since the best results were obtained using 10 principal components, a single FFDM image took an average of 185 seconds (approximately 3 minutes). FFDM with large breast areas were observed taking as long as 6 minutes to complete, while FFDM images with very little breast area were able to complete in just under a minute.

5.4 Accuracy of Computed Breast Density

The results of Oliver's implementation show a 94% accuracy rate for pixel classification. However, the accuracy rate on a per-image basis is not available due to the fact that the images being used in the study differ from the ones used by Oliver. An assumption is made that the per-image accuracy rate would be significantly high, especially when taking into account the inter-observer variance discussed earlier.

For the purposes of this study, accuracy was defined as any density value

computed from an FFDM image which is within the industry accepted inter-observer error of the ground truth value. An inter-observer error value of 20% was chosen as it was the lower bound of the accepted error ranges between any two trained radiologists.

Plotting the algorithm's density values for each FFDM against the corresponding ground truth values, as shown in Figure 5.8, shows an extremely strong correlation between computed density and ground truth observed density which assists in validating the results published by Oliver.

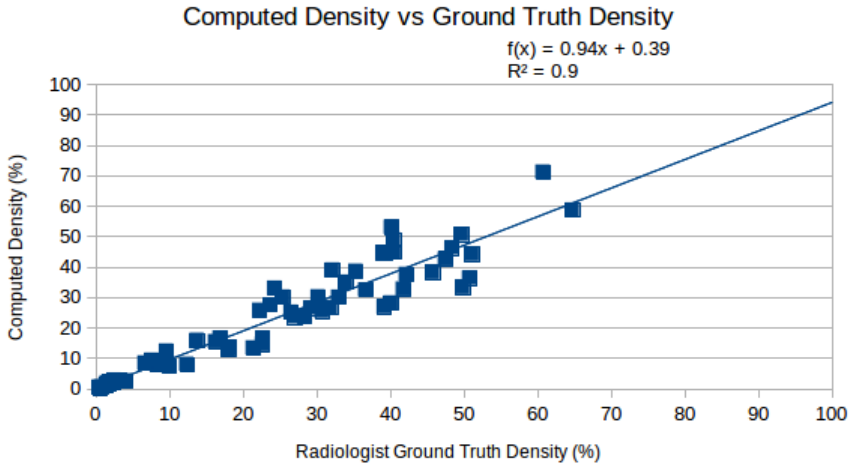


Figure 5.8: Computed density vs. ground truth.

Using linear regression, the slope of the line of best fit indicates that, overall, the algorithm presents a slight bias toward under-estimating breast density. A strong linear correlation was an expected result since the algorithm is supervised and the training images densities were classified by the same radiologists used in generating the ground truth. Additionally, Figure 5.8 shows an increase in output variation as breast density increases;

mimicking actual radiologist observational data.

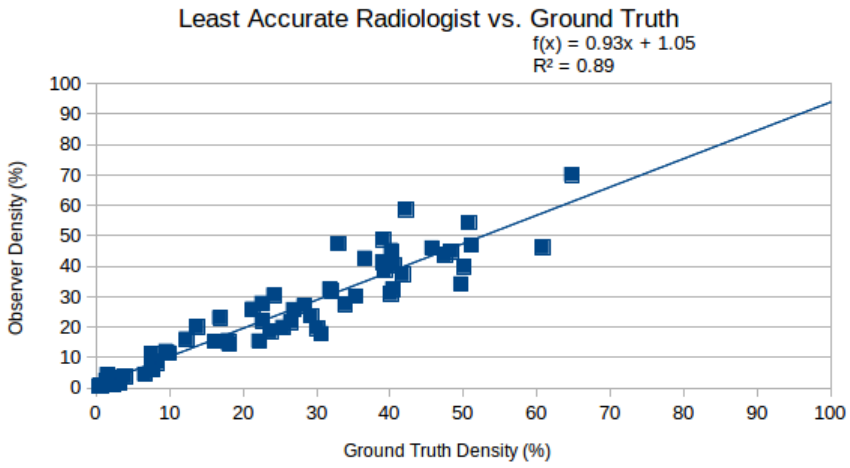


Figure 5.9: Observer 1 density vs. ground truth.

To compare the overall performance of the algorithm to that of a trained radiologist, Figure 5.9 plots an individual participating radiologist's observations against the ground truth values. The radiologist in question was the least accurate from the set of four, with an R^2 of 0.89. The most accurate radiologist R^2 value was 0.92. These values place the algorithm's performance within the measured inter-observer range of correlation and, when considering that the four participating radiologists had an average absolute error of only 11.5%, well within the expected inter-observer error between two trained radiologists.

The error rates for the FFDM image set used in this study are shown in Figure 5.10. Error rates were computed using the following equation:

$$E_{\%} = \left(1.0 - \frac{|D_{rad} - D_{computed}|}{D_{rad}}\right) \times 100 \tag{5.1}$$

where D_{rad} and $D_{computed}$ are the density values determined by the radiologist ground truth and density algorithm, respectively. Overall, the algorithm implementation scored perfectly when using an error tolerance of 20% which is similar to the expected inter-observer error between experienced radiologists. As error tolerance was lowered, the error increased. At a tolerance of 5% error, the accuracy was below the accepted value of radiologist inter-observer error.

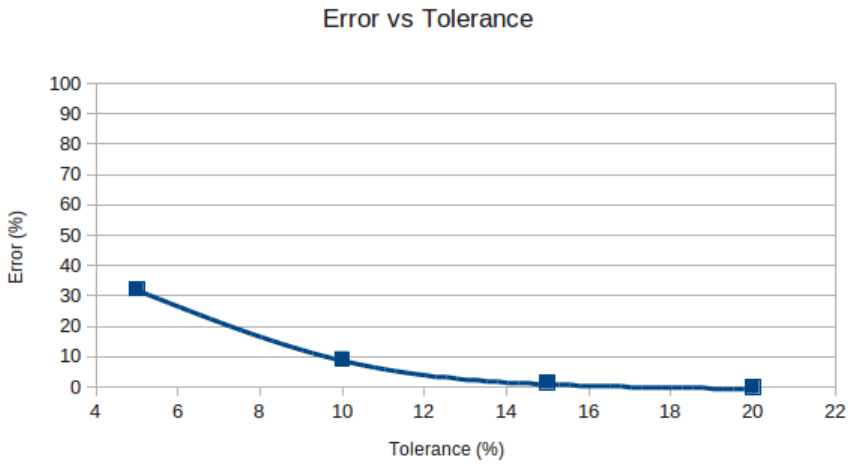


Figure 5.10: Error vs. Tolerance.

5.5 Size of Classification Region

In Oliver's algorithm, the overall breast density was determined by visiting each pixel inside the breast area of an FFDM and making a dense or fatty classification for that particular pixel. The algorithm's elapsed time for this process was, on average, 3 minutes. To explore the effects of decreasing the algorithm's running time versus accuracy, the pixel-by-pixel approach was substituted for a region-by-region method giving the output binary image less overall resolution, as shown in Figure 5.11.

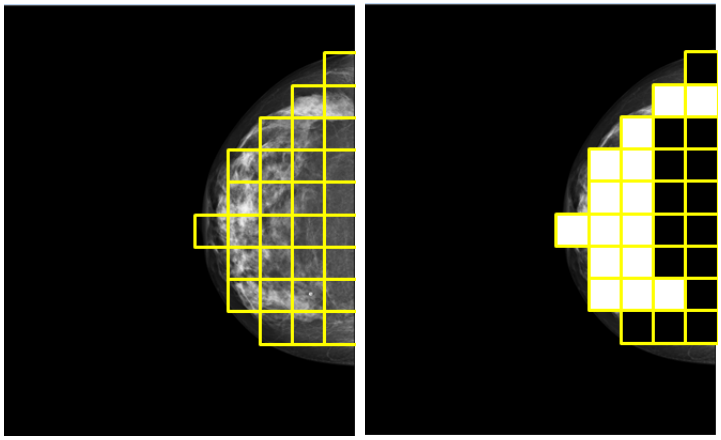
(a) Output from original implementation

(b) Output from modified implementation

Figure 5.11: Comparison of outputs from algorithm modification.

Instead of visiting each pixel within the breast area, the algorithm was modified to divide the breast area into a grid, visit each sub-region, and finally centre the 50×50 classification window on the current sub-region to perform analysis as shown in Figure 5.5. Instead of classifying the individual pixel, the entire sub-region was classified. The 50×50 classification remained constant throughout giving an upper limit to the allowable size of the sub-

region.



(a) The breast is now examined region-by-region

(b) Each region is classified as dense or fatty.

Figure 5.12: Region-by-region classification.

Figure 5.14 shows the effect on accuracy of substituting the pixel-by-pixel approach for a region-by-region approach. The drop in correlation between the computed breast density and the ground truth is, in fact, quite negligible with a difference in R^2 value of less than 0.02. Translated into variation of computed density, the loss of correlation amounts to no more than a maximum additional 2% error in computed density values for a few of the FFDM images when using a region size of 50×50 as shown in Figure 5.13.

While a potential 2% increase in error may be undesirable, it may be offset by the dramatic decrease in algorithm running time. The total FFDM image set took approximately 3.3 hours to complete using the pixel-by-pixel method, however, the region-by-region approach running time was reduced to a mere 5 seconds.

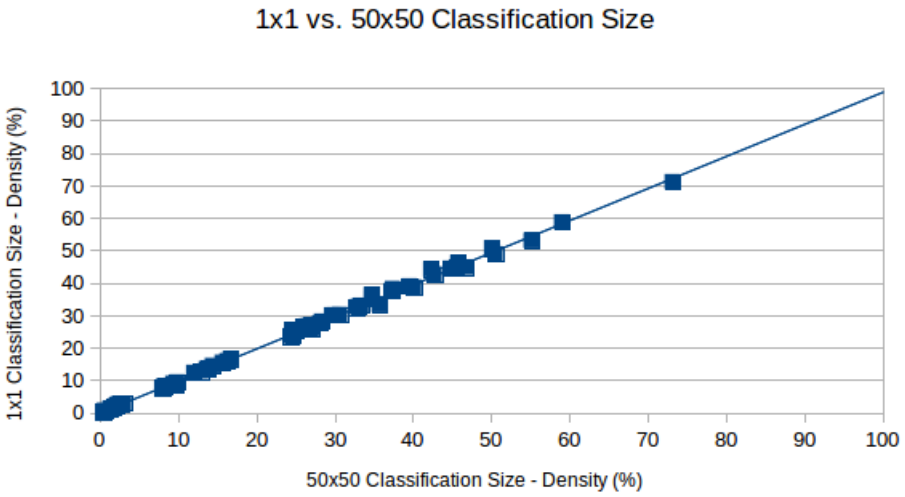


Figure 5.13: Comparison of min and max classification sizes.

The optimal trade-off between accuracy and processing time was found at the intersection of Figures 5.14 and 5.15. A region size of 10×10 pixels decreases the running time by a factor of 95 while giving a maximum increase in error of 0.39%.

5.6 Effect of Image Bit-Depth

Since the motivation for developing automated breast density computation algorithms comes from its strong correlation with breast cancer, long-term storage of FFDM images is necessary to amass a large collection of research data. With the increased number of FFDM images potentially being stored comes an increase in associated storage costs. To help reduce these storage costs, methods of minimizing the required storage per FFDM images could be implemented. However, the implementation of storage com-

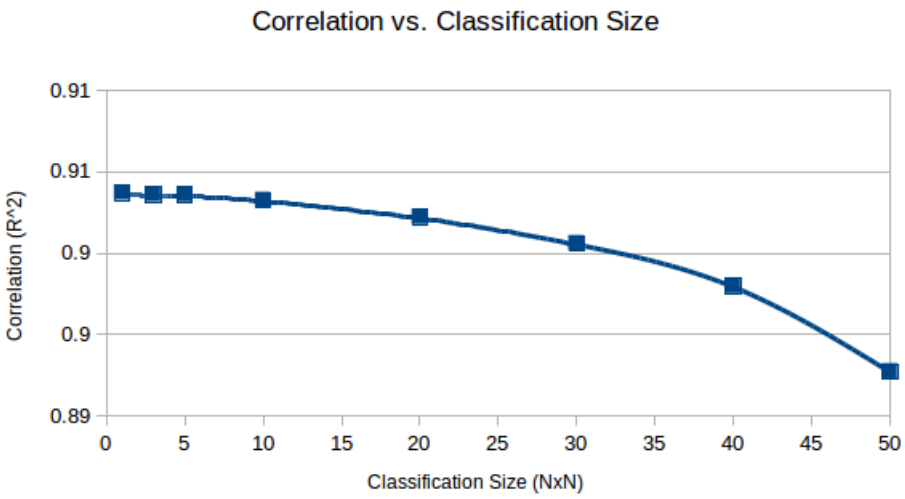


Figure 5.14: Effect of classification size on algorithm accuracy.

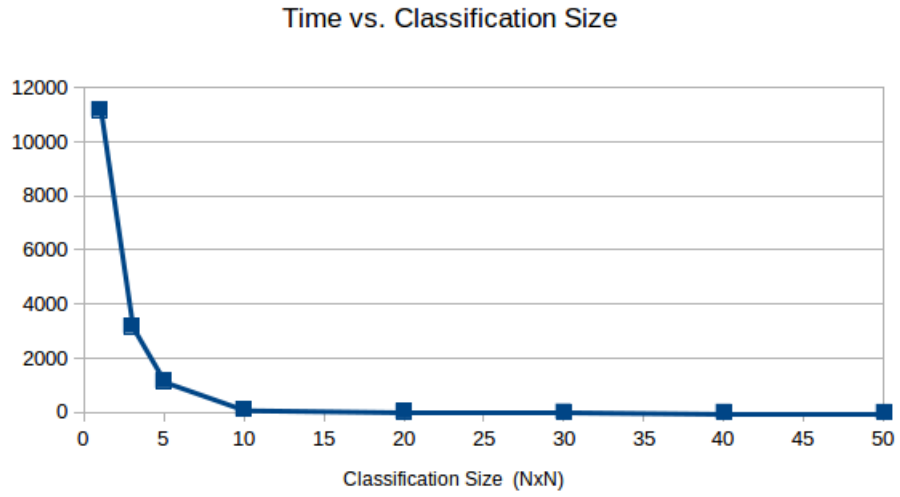


Figure 5.15: Effect of classification size performance.

pression usually results in a loss of image data and/or image contrast. To determine the effects of compression techniques, the algorithm was re-run on a reduced-contrast version of the FFDM image data set. FFDM images were compressed to contain no more than 8-bits of gray-scale storage per pixel; down from the original 12-bits used in standard diagnostic FFDM images. The reduction in required storage for 8-bit FFDM images was 4 times less than what was originally required. This was due to the fact that even though the standard FFDM images used a maximum of 12-bits per pixel, 16-bits of storage was still allocated per pixel for convenience.

As shown in Figure 5.16, the computed densities for the 8-bit versions of the FFDM images are virtually identical to values computed from the standard FFDM image set. The average difference in computed densities between the two image sets was less than 0.022% with a standard deviation of 0.02%.

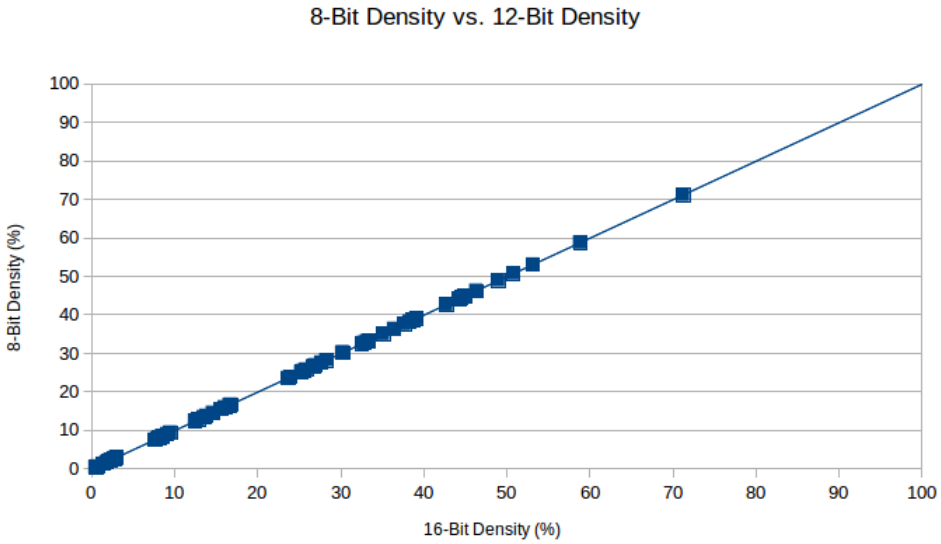


Figure 5.16: Effect of image colour depth.

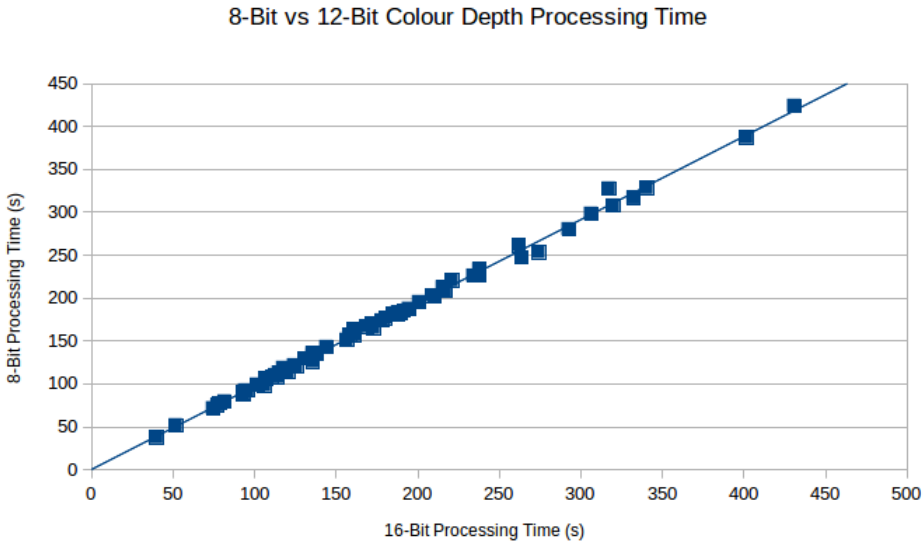


Figure 5.17: Effect of image colour depth on performance.

The algorithm's processing time was only negligibly affected by the substitution of 8-bit FFDM images, as shown in Figure 5.17. This result was expected since modern computer processor caching hardware and memory addressing is optimized for 32-bit values. Images with 8-bits were, on average, 4.82 seconds faster than their 12-bit counterparts. This could equate to an average 2.8% reduction in algorithm processing time throughout the entire 8-bit image set. However, the standard deviation of 4.84 seconds and average error of 4.82 seconds shows that the savings are not guaranteed on a per-image basis. In fact, a very small portion of the 8-bit FFDM images actually ran slightly slower compared to the 12-bit version of the same image.

5.7 Effect of Raw Image Data

As mentioned in Section 2.1, FFDM images are produced via the application of proprietary image enhancement techniques on the raw sensor data. During the application of these techniques, a large portion of image contrast is lost and therefore unavailable for analysis. While the raw sensor data is never used for diagnostic purposes or stored with a patient's electronic medical record, it can be retrieved directly from the digital mammography device for a short period. With access to the corresponding raw sensor data images from our FFDM image data set, the algorithm's performance on high-contrast raw FFDM images was investigated.

Raw FFDM images contain 14-bit gray-scale values per pixel as opposed to the 12-bits used in diagnostic FFDM images. In Figure 5.18, the com-

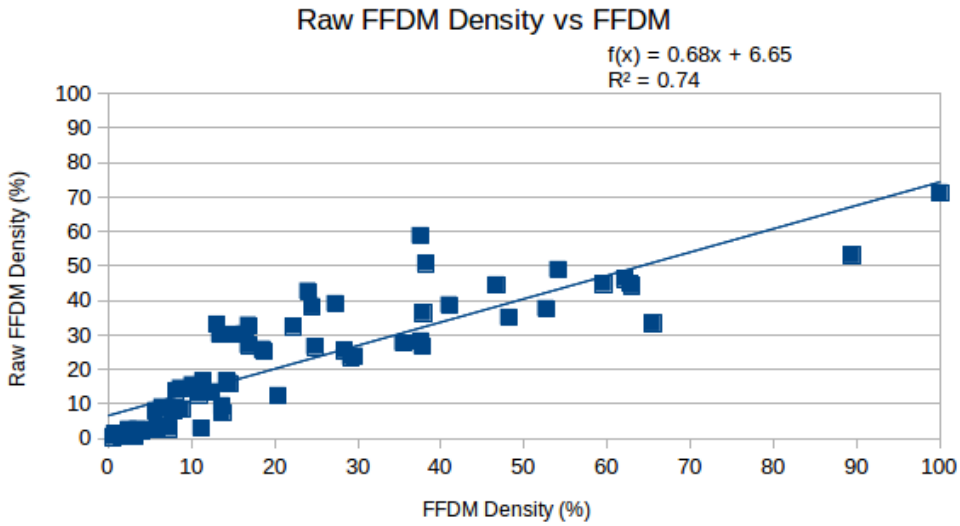


Figure 5.18: Effect of using raw images.

puted breast density values for the raw FFDM images were plotted against the density values obtained from the regular diagnostic FFDM images. The R^2 correlation between the two data series was 0.74 showing a loose correlation between the two series. Also of note was the tendency for the raw FFDM images to be slightly higher than densities obtained with diagnostic FFDM images. To ensure that the training criteria was identical for both the raw, and diagnostic FFDM comparisons, the training images for the raw FFDM analysis were created using the identical spatial coordinates as training images created for the diagnostic FFDMs. The spatial coordinates for both the raw and diagnostic FFDM images correspond to the identical pixel data, regardless of image optimization techniques used.

This result was surprising since, fundamentally, there was very little difference between the approach used on diagnostic FFDM images and the

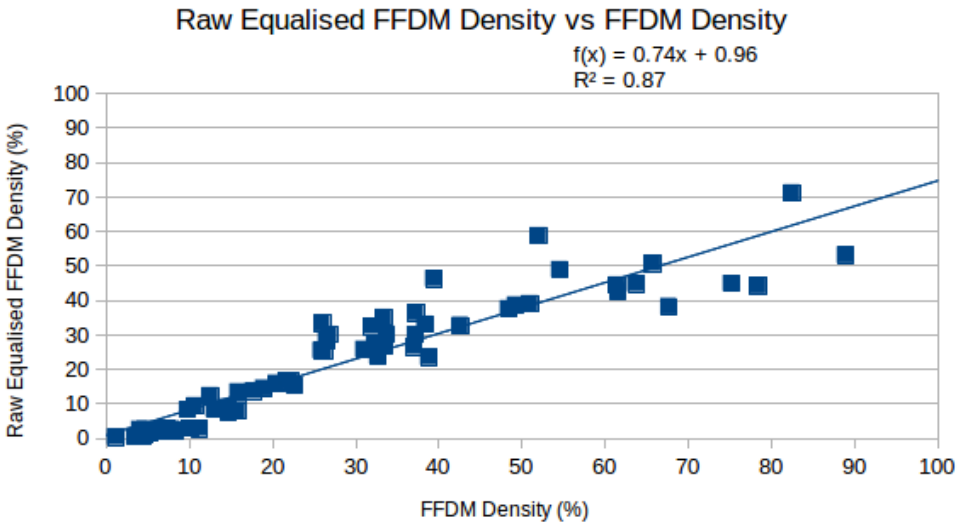


Figure 5.19: Effect of using equalised raw image.

raw sensor data images. To investigate further, the raw images were examined more closely; specifically the region of the images containing the breast tissue. The investigation revealed that while the overall image contrast in raw FFDM images is much higher than the diagnostic FFDM images, the contrast range within areas of breast tissue were actually much lower; hence the reasoning for image processing to produce diagnostic FFDM images. To further examine this hypothesis, the raw images underwent histogram equalization after the breast tissue was segmented from the background pixels. The algorithm was then re-run.

Figure 5.19 shows the results of density values computed with the raw FFDM images which underwent histogram equalization against the unaltered raw FFDM images. While there is a slight tendency for the equalised raw FFDM images to be lower in density value, the R^2 value increased to

0.87 indicating that histogram equalization is a major contributing factor when applying the Eigenfaces algorithm to FFDM images.

5.8 BI-RADS Classification

In medical practice, the quantification of breast density is usually performed through BI-RADS classification to help minimize inter-observer error and establish an agreed-upon set of criteria for fibroglanduar tissue density. Of interest was this algorithm's performance when asked to classify each FFDM image into its corresponding BI-RADS category.

During the initial recordings of observations from the participating radiologists, both the percentage of breast density values and BI-RADS classifications were obtained. To determine the algorithm's performance, the relationship between the recorded density values and BI-RADS classifications were studied.

From the sample set of images used during this experiment, BI-RADS category I images were consistently rated as those with a density percentage between 0% and 5%. Images with a BI-RADS II and III categorisation contained densities between 5% to 30% and 20% to 50%, respectively. BI-RADS IV contained those images having densities exceeding 50%. Figure 5.20 shows a histogram containing the number of images within each density range and their corresponding BI-RADS classification.

To determine classification performance, the algorithm was modified to select a BI-RADS category for each of the FFDM images within the data set. The algorithm chose the appropriate BI-RADS category by analyzing

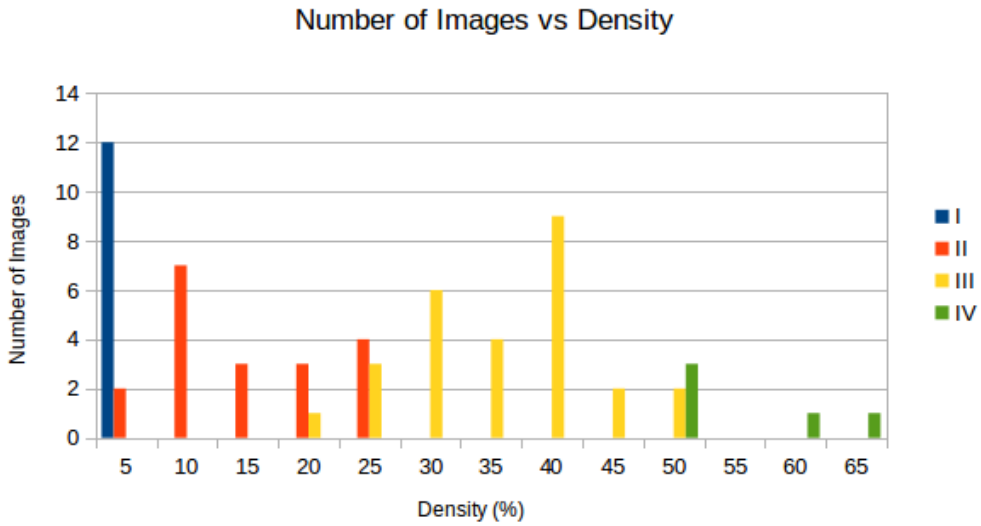


Figure 5.20: Histogram of BI-RADS classifications.

the probability distribution histogram which was prepared in advance by computing the average density within each BI-RADS category and plotting a Guassian distribution from the average out to 1 standard deviation. Once the density of an FFDM image was computed, the algorithm assigned the BI-RADS category which contained the highest probability at the computed density. The data used for comparison is shown in Figure 5.21.

Table 5.2: Algorithm BI-RADS classification accuracy.

BI-RADS Category	*Accuracy*
I	100%
II	84.2%
III	77.7%
IV	100%
Overall	84.6%

Table 5.2 shows the accuracy rates for the algorithm's BI-RADS clas-

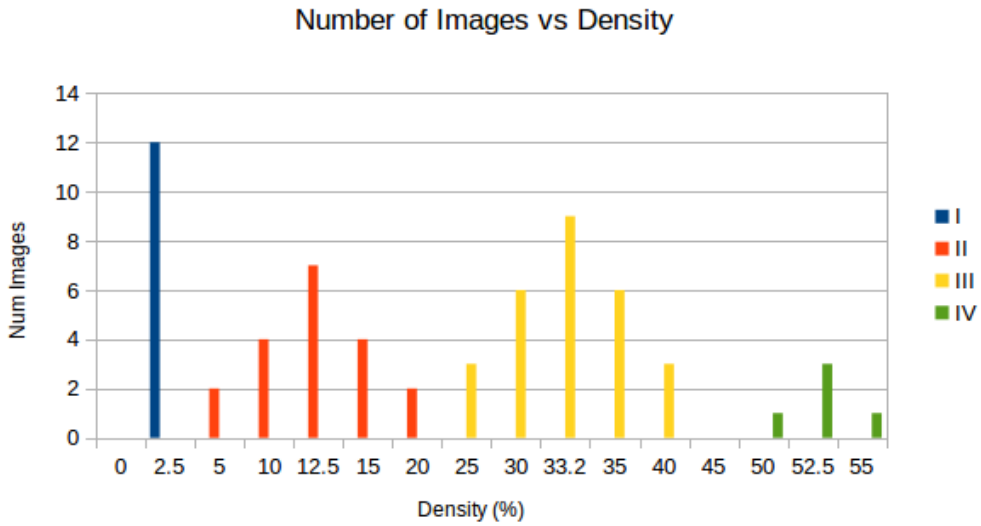


Figure 5.21: Probability of BI-RADS classification.

sification for each corresponding categories. For extremely fatty breasts, the algorithm was 100% accurate. Given that such mammogram images are easy to detect and their densities are consistently at or below 5%, such a result was to be expected. Similarly, high-density breasts were also categorised with 100% accuracy due to their pronounced computed density values. However, a true 100% accuracy rate for both fatty and highly dense breasts is unlikely since the population sample size used for this experiment contained only 12 and 5 images, respectively.

For BI-RADS categories II and III, the accuracy rate dropped to between 77.7% and 84.2%, respectively. For images where its density was well within the probability range of a known category, the accuracy rate was extremely high, however, accuracy rates declined substantially when an image's density fell between two categories. Using the closest category method

for determining BI-RADS classification proved to only slight more effective than chance.

Chapter 6

Conclusion

From the results obtained, an automated breast density algorithm can be loaded with high-quality training images and achieve results which are not only consistent, but correlate strongly with the observations from trained and experienced radiologists. Moreover, the algorithm's accuracy rate from the ground truth observations was shown to be extremely accurate when compared against the currently accepted inter-observer error rate of 20-30%. When compared against four highly trained and experienced radiologists, the algorithm's accuracy was consistently accurate to tolerances of 10% of the ground truth value and was shown to correlate at least as well as the least accurate participating radiologist.

The number of training images used to calibrate the algorithm was shown to dramatically affect accuracy if an insufficient quantity is utilised. An optimal number of training images for this investigation was found to be 120 comprised of 60 dense and 60 fatty images. Adding an additional quantity of training images beyond this value did not add a measurable benefit to the algorithm's accuracy.

The algorithm's accuracy is influenced by the quality of the training images used to identify unknown FFDM images. Training images which

were taken from extremely fatty and dense FFDM images gave optimal results compared to training images obtained randomly or whose densities were close in value. Deviations from the optimal training image set reduced the R^2 value by as much as 3%. Poor quality training images, or training images which were improperly classified had a detrimental effect on the algorithm's overall accuracy.

Speed and performance of the algorithm were also measured and explored. To ensure optimal accuracy, the number of principal components used during the PCA transform was examined experimentally. The results indicate that using more than 10 principal components yields a negligible increase in overall algorithm accuracy. Additionally, the use of more than 10 principal components increases the running time of the algorithm logarithmically without any performance contribution.

To improve further the algorithm's running time performance from its original implementation, the optimal number of pixels used to denote dense tissue was explored. In its original form, the algorithm examined an FFDM image with unknown density on a pixel-by-pixel basis. Each pixel was examined individually and then classified as either dense or fatty. To decrease running time, the algorithm was modified to examine only the pixel centred within a square region. The entire square region was then classified as dense of fatty depending upon how the centre pixel was classified. The decrease in running time was dramatic as the square region increased in size. Running time decreased from 4.3 minutes per image to just a few seconds by the time the region reached a dimension of 10×10 pixels. As the region continued to expand, the change in running time became much less significant, however

the algorithm accuracy was not significantly altered; even when the region grew to the maximum size of 50×50. At maximum size, the accuracy diminished by a maximum of 2%. While maximum accuracy is the desired trait in any algorithm, running time can be an expensive commodity during real-world implementations. These results indicate that running time can be substantially decreased with only minor impairment of accuracy. It is important to note, however, that improvements in running time without any loss in accuracy could also be achieved by implementing a parallelized version of the algorithm.

Also of concern during any real-world implementation is the amount of digital storage required. While clinical systems are bound by legislation regarding image type and resolution which must be kept on record, images for academic use are not. To determine the effects of storage compression techniques, the algorithm was modified to accept FFDM images with a decreased bit-depth; 8-bits instead of the usual 12-bits. Since images must use a bit-depth which is a multiple of 8 bits, 12-bit images are, in fact, stored as 16-bit images with the upper 4-bits unused. 8-bit image utilise the entire 8-bits and decrease the amount of required digital storage by 50%. This comes at the expense of lost image information, which while often undetectable by human eyes, can have a significant impact on image processing algorithms. However, the algorithm performed almost identically when 8-bit FFDM images were substituted for the 12-bit images. Differences in the computed densities were measurable, but negligible, and running time was also not affected. While this finding may not yield benefits for clinical implementations, academic researchers wishing to store large data sets with

limited budgets may find savings in storage requirements.

Investigation into the use of raw FFDM images substituted for diagnostic-grade FFDM images showed that the density algorithm is not able to directly convert input image types. This was an unexpected result since the supervised image analysis techniques and principles can be applied with either image set. However, while raw images contain more gray-level variation compared with their respective diagnostic-grade counterparts, the gray-level contrast for breast tissue is contained within only a very small range. To determine if this fact played a significant role in the density algorithm's decreased performance, the raw images were altered via an image processing technique known as histogram equalization which maximises the use of the available gray-scale range. Results from this altered raw image set more closely aligned with the results obtained with the diagnostic-grade image set, however, some performance degradation was still observed.

In conclusion, the algorithm has been shown to be as good at quantifying and classifying breast density as the least accurate radiologist in this study. Moreover, the inter-observer error observed from the algorithm easily falls within the accepted error of trained radiologists. It has also been shown that the algorithm's running time can be significantly improved while suffering very little degradation in accuracy.

6.1 Future Work

Additional research is required to achieve greater performance with existing automated breast density image processing algorithms. From the results

obtained, it is clear that the selection of the training images and patches has an impact on the overall accuracy of the density algorithm. In Oliver's research, an optimal quantity and variation of training data is briefly discussed with reference to its discovery through empirical methods. It is extremely likely that the training data which Oliver found to be optimal for his algorithm implementation is a locally optimized one, and not globally optimal. More investigation into the characteristics of optimal training data is needed to ensure a globally optimal set can be obtained from an extremely large sample.

During the phase of the algorithm where principal component analysis is utilised, an unknown patch is classified by the training patch which has the smallest Euclidean distance within the PCA subspace. Other methods for classification with clustered data are also known to exist, such as the Mahalanobis distance [21]. The effects on algorithm performance with an alternative distance quantification technique is unknown at this time and warrants further investigation.

The core inspiration for the original implementation of Oliver's breast density algorithm is the Eigenfaces facial recognition algorithm. While this recognition algorithm has been widely implemented for use in many image processing applications with great success, its use as a density classification engine is largely unexplored. Additionally, the Eigenfaces algorithm is relatively dated and more modern classification engines, such as Laplacianfaces [10] have been introduced as a possible successor. The use of such a recognition algorithm may lead to increased performance as well as a reduction in the quantity of training data required.

Since other image processing techniques for determining breast density have been successful in the past, the effect of combining multiple techniques in conjunction with Oliver's density algorithm could produce an increase in density accuracy. Of notable interest is the technique of histogram classification which has previously been applied to the domain of breast density with great success [36] when using analog mammogram images. Since Oliver's implementation shows the greatest inaccuracy when analyzing images within the BI-RADS II and III categories, histogram classification could boost performance by adding an external reference for images with specific histogram profiles.

To further increase performance, histogram analysis could assist Oliver's implementation by providing an ability to detect FFDM images with extremely low or high breast density. The profiles of such images have histogram profiles which are easily recognisable by simply computing the range of contrast.

With additional research and improvements, it is likely that automated breast density classification algorithms will soon surpass even the most seasoned radiologists in terms of inter-observer accuracy. However, most importantly, these algorithms will provide an extremely consistent method for quantifying breast cancer risk and possibly help establish an industry standard for breast density measurement.

Bibliography

[1] P. Bakic, A. Carton, D. Kontos, C. Zhang, A. Troxel, and A. Maidment, *Breast Percent Density: Estimation on Digital Mammograms and Central Tomosynthesis Projections1*, Radiology, 252 (2009), pp. 40–49. → pages 8

[2] W. E. Barlow, C. Chi, P. a. Carney, S. H. Taplin, C. D'Orsi, G. Cutter, R. E. Hendrick, and J. G. Elmore, *Accuracy of screening mammography interpretation by characteristics of radiologists.*, Journal of the National Cancer Institute, 96 (2004), pp. 1840–50. → pages 8

[3] J. C. Bezdek, *FCM : The Fuzzy C-means Clustering Algorithm*, Computers & Geosciences, 10 (1984), pp. 191–203. → pages 14

[4] J. Boone, K. Lindfors, and C. Beatty, *A breast density index for digital mammograms based on radiologists' ranking*, Journal of Digital Imaging, 13 (1998), pp. 1–2. → pages 11

[5] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. a. Jong, G. Hislop, A. Chiarelli, S. Minkin, and M. J. Yaffe, *Mammographic density and the risk and detection of breast*

cancer., The New England journal of medicine, 356 (2007), pp. 227–36. → pages 25

[6] J. Byng, N. Boyd, and E. Fishell, *Automated analysis of mammographic densities*, Physics in medicine, 909 (1999). → pages 10

[7] S. Ciatto, N. Houssami, a. Apruzzese, E. Bassetti, B. Brancato, F. Carozzi, S. Catarzi, M. P. Lamberini, G. Marcelli, R. Pellizzoni, B. Pesce, G. Risso, F. Russo, and a. Scorsolini, *Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories.*, Breast (Edinburgh, Scotland), 14 (2005), pp. 269–75. → pages 8

[8] I. T. Gram, Y. Bremnes, G. Ursin, G. Maskarinec, N. Bjurstam, and E. Lund, *Percentage density, Wolfe's and Tabár's mammographic patterns: agreement and association with risk factors for breast cancer.*, Breast cancer research : BCR, 7 (2005), pp. R854–61. → pages 1

[9] C. L. Hart and G. Erbacher, *Mammogram interpretation using the BI-RADS final assessment categories in 40- to 49-year-old women*, Journal of the American Osteopathic Association, 100 (2000), pp. 615–619. → pages 8

[10] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, *Face recognition using laplacianfaces*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27 (2005), pp. 328–340. → pages 58

[11] J. J. Heine, K. Cao, and D. E. Rollison, *Calibrated measures for breast density estimation.*, Academic radiology, 18 (2011), pp. 547–55. → pages 8, 11

[12] J. J. Heine, K. Cao, D. E. Rollison, G. Tiffenberg, and J. a. Thomas, *A quantitative description of the percentage of breast density measurement using full-field digital mammography.*, Academic radiology, 18 (2011), pp. 556–64. → pages 11

[13] J. J. Heine and R. P. Velthuizen, *A statistical methodology for mammographic density detection.*, Medical physics, 27 (2000), pp. 2644–51. → pages 10

[14] K. Kerlikowske, D. Grady, J. Barclay, S. D. Frankel, S. H. Ominsky, E. a. Sickles, and V. Ernster, *Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System.*, Journal of the National Cancer Institute, 90 (1998), pp. 1801–9. → pages 7, 8

[15] J. Kittler and J. Illingworth, *Minimum error thresholding*, Pattern Recognition, 19 (1986), pp. 41 – 47. → pages 10

[16] C. I. Lee, L. W. Bassett, and C. D. Lehman, *Breast density legislation and opportunities for patient-centered outcomes research*, Radiology, 264 (2012), pp. 632–636. → pages 1

[17] S. Lloyd, *Least squares quantization in PCM*, IEEE Transactions on Information Theory, 28 (1982), pp. 129–137. → pages 13

[18] L. Lu, T. K. Nishino, T. Khamapirad, J. J. Grady, M. H. Leonard, and D. G. Brunder, *Computing mammographic density from a multiple regression model constructed with image-acquisition parameters from a full-field digital mammographic unit.*, Physics in medicine and biology, 52 (2007), pp. 4905–21. → pages 10

[19] B. Lundgren and S. Jakobsson, *Single view mammogrphy: A simple and efficient approach to breast cancer screening*, Cancer, 38 (1976), pp. 1124–1129. → pages 3

[20] P. J. Lynch, *Breast anatomy normal scheme.png*, 2007. → pages vii, 4

[21] R. D. Maesschalck, D. Jouan-Rimbaud, and D. Massart, *The mahalanobis distance*, Chemometrics and Intelligent Laboratory Systems, 50 (2000), pp. 1 – 18. → pages 58

[22] P. of Canada, *Private member's bill c-314*, 2013. → pages 1

[23] A. Oliver, J. Freixenet, R. Marti, J. Pont, E. Pérez, E. Denton, and R. Zwiggelaar, *A novel breast tissue density classification methodology*, Information Technology in Biomedicine, IEEE Transactions on, 12 (2008), pp. 55–65. → pages 14

[24] A. Oliver, J. Freixenet, R. Martí, and R. Zwiggelaar, *A comparison of breast tissue classification techniques.*, Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 9 (2006), pp. 872–9. → pages 14

[25] A. Oliver, X. Llado, R. Marti, and J. Freixenet, *Classifying mammograms using texture information*, Medical Imaging, (2007), pp. 223–227. → pages 27

[26] A. Oliver, X. Lladó, E. Pérez, J. Pont, E. R. E. Denton, J. Freixenet, and J. Martí, *A statistical approach for breast density segmentation.*, Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology, 23 (2010), pp. 527–37. → pages 14

[27] E. Ooms, H. Zonderland, M. Eijkemans, M. Kriege, B. M. Delavary, C. Burger, and A. Ansink, *Mammography: Interobserver variability in breast density assessment*, The Breast, 16 (2007), pp. 568 – 576. → pages 8

[28] R. Rajapakshe, B. Uyaniker, P. Gordon, and S. Silver, *A fully automatic method for estimating breast density in digital mammograms*, Radiological Society of North America Scientific Assembly and Annual Meeting Program, Oakbrook Ill., SSM01-03 (2009), p. 574. → pages 8

[29] P. Saha, J. Udupa, E. Conant, D. Chakraborty, and D. Sullivan, *Breast tissue density quantification via digitized mammograms*, Medical Imaging, IEEE Transactions on, 20 (2001), pp. 792–803. → pages 10

[30] N. Saidin, U. Ngah, and H. Sakim, *Density based breast segmentation for mammograms using graph cut techniques*, TENCON 2009-2009, (2009), pp. 1–5. → pages 8

[31] R. Sivaramakrishna, N. a. Obuchowski, W. a. Chilcote, and K. a. Powell, *Automatic segmentation of mammographic density.*, Academic radiology, 8 (2001), pp. 250–6. → pages 10

[32] F. Squires, J. R. Glassman, A. Morris, and O. Thepurpose, *The Breast Imaging Reporting and Data System : Positive Predictive Value of Mammographic Features and Final AssessmentCategories*, American journal of roentgenology, 73 (1998), pp. 35–40. → pages 7

[33] A. Torrent, A. Bardera, A. Oliver, and J. Freixenet, *Breast density segmentation: a comparison of clustering and region based techniques*, in Digital Mammography, 2008, pp. 9–16. → pages 8

[34] A. Turk, M.; Pentland, *Eigenfaces for Recognition*, Journal of cognitive neuroscience, 3 (1991), pp. 71–86. → pages 14

[35] J. Wolfe, *Risk for breast cancer development determined by mammographic parenchymal pattern*, Cancer, 37 (1976), pp. 2486–2492. → pages 1

[36] C. Zhou, C. Heang-Ping, N. Petrick, B. Sahiner, and et. al., *Computerized image analysis: estimation of breast density on mammograms*, (2000), pp. 1615–1624. → pages 1, 10, 59

Appendix

Appendix A

Main Program

```
int
main(int argc, char **argv)
{
  cv::Mat patch_img;
  map<string, int> patches;
  vector<cv::Mat> patchImgs;
  vector<int> labels;
  size_t patch_sz;
  Ptr<FaceRecognizer> model;
  string patch_fname;
  Image<BD_PIXEL_TYPE> breast;
  Image<BD_PIXEL_TYPE> out;
  Image<BD_PIXEL_TYPE> mask;
  Image<BD_PIXEL_TYPE> *trimg;
  size_t dense_pixels;
  size_t breast_pixels;
  Pixel<BD_PIXEL_TYPE> pixel;
  Region patch;
```

```
int classification;
float density;
double algoseconds;
string breast_fname;
string mask_fname;
string out_fname;
int windowsz;
int num_pca;
const char *output_dir, *patch_dir;

parse_cmd_args(argc, argv, &windowsz,
               &num_pca, &output_dir,
                                 &patch_dir);

if (patch_dir == NULL)
  {
    cerr <<
             "please specify the patch directory"
             << endl;
    exit(EXIT_FAILURE);
  }

if (output_dir == NULL)
  {
```

```
    cerr <<
            "please specify the output directory"
            << endl;
    exit(EXIT_FAILURE);
  }

breast_fname = argv[argc - 2];
mask_fname = argv[argc - 1];

patch_sz = read_patchdb(PATCHDB_FILE, patches);

for (map<string, int>::const_iterator i = patches.begin();
    i != patches.end();
    i++)
  {
    pair<string, int> p = *i;

    patch_fname = patch_dir;
    patch_fname.append("/");
    patch_fname.append(p.first);

    patch_img = imread(patch_fname, -1);
    if (patch_img.data == NULL)
      {
```

```
            cerr << "unable to load patch image '"
                        << patch_fname << "'" << endl;
            exit(EXIT_FAILURE);
          }

        patchImgs.push_back(patch_img);
        labels.push_back(p.second);
      }

    model = createEigenFaceRecognizer(num_pca);
    model->train(patchImgs, labels);

    try
      {
        breast.loadFile(breast_fname);
      }
    catch (std::exception& e)
      {
        cerr << "unable to open breast image '"
                  << breast_fname << "'" << endl;
        cerr << e.what() << endl;
        exit(EXIT_FAILURE);
      }
```

```
try
  {
    mask.loadFile(mask_fname);
  }
catch (std::exception& e)
  {
    cerr << "unable to open mask image '"
         << mask_fname << "'" << endl;
    cerr << e.what() << endl;
    exit(EXIT_FAILURE);
  }

density = 0;

if (windowsz == 0)
  density =
        ComputeBreastDensity<BD_PIXEL_TYPE>(breast,
        mask, out, patch_sz, model, &algoseconds);
else
  density =
        ComputeBreastDensity<BD_PIXEL_TYPE>(breast,
        mask, out, windowsz, patch_sz, model, &algoseconds);

cout << BaseName(breast_fname) << ",";
```

```
    cout << (density * 100.0) << ",";
    cout << algoseconds << endl;

    // Write the results
    out_fname = output_dir;
    out_fname.append("/");
    out_fname.append(BaseName(breast_fname));

    try
    {
        out.saveFile(out_fname.c_str());
    }
    catch (std::exception& e)
    {
        cerr << "unable to save file '"
                << out_fname << "'" << endl;
        cerr << e.what() << endl;
        exit(EXIT_FAILURE);
    }

    return EXIT_SUCCESS;
}
```

Appendix B

Density Algorithm

```
template<class T>
  float
  ComputeBreastDensity(
          const Image<T>& breast,
          const Image<T>& mask,
      Image<T>& out, size_t patchsz,
          cv::Ptr<cv::FaceRecognizer>& model,
      double *seconds)
  {
    size_t dense_pixels;
    size_t breast_pixels;
    Image<T> work;
    Image<T> *trimg;
    cv::Mat m;
    Pixel<T> pixel;
    Region patch;
    int classification;
    struct timeval start, stop, diff;
```

```
breast.copyTo(work);

work.applyMask(mask);
work.copyTo(out);

out.setAllPixels(0);

// Note the time of algorithm start.
gettimeofday(&start, NULL);

// Beginning pixel-by-pixel algorithm.
for (size_t y = patchsz / 2;
         y < work.getRows() - patchsz / 2;
             y++)
  {
    for (size_t x = patchsz / 2;
                 x < work.getCols() - patchsz / 2;
                     x++)
      {
        pixel = work.getPixel(bd::Point(x, y));

        if (pixel.value == 0)
          continue;
```

```
patch = Region::createOnCentre(x,
                y, patchsz, patchsz);

if (mask.hasZerosInRegion(patch))
  continue;

trimg = new Image<T>(patchsz, patchsz);
work.createTrainingImage(patch, *trimg);

if (sizeof(T) == 1)
  m = cv::Mat(trimg->getRows(),
                    trimg->getCols(),
                    CV_8UC1,
  (void *) trimg->getPixels());
else if (sizeof(T) == 2)
  m = cv::Mat(trimg->getRows(),
                    trimg->getCols(),
                    CV_16UC1,
  (void *) trimg->getPixels());
else
  throw std::invalid_argument(
                    "image type not supported"
                    );
```

```
            classification = model->predict(m);

            if (classification == 1)
              out.setPixel(
                              Pixel<T>(x, y,
                              BD_PIXEL_MAX_VALUE)
                                  );
            else
              out.setPixel(Pixel<T>(x, y, 0));

            delete trimg;
          }
        }

    // Note time of algorithm stop.
    gettimeofday(&stop, NULL);

    TimevalDiff(&diff, &stop, &start);

    *seconds = diff.tv_sec +
                    (diff.tv_usec / 1000000.0);

    dense_pixels = out.getNumPixelsAboveThreshold(0);
```

```
    breast_pixels = mask.getNumPixelsAboveThreshold(0);

    return dense_pixels / (float) breast_pixels;
  }
```