UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPING NOVEL COMPUTER-AIDED DETECTION AND DIAGNOSIS SYSTEMS OF MEDICAL IMAGES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

FARANAK AGHAEI
Norman, Oklahoma
2019

DEVELOPING NOVEL COMPUTER-AIDED DETECTION AND DIAGNOSIS

SYSTEMS OF MEDICAL IMAGES

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Bin Zheng, Chair

Dr. Hong Liu

Dr. Joseph P. Havlicek

Dr. Liangzhong Xiang

Dr. Tomasz Przebinda

Dr. Han Yuan

Acknowledgements

Many people helped or played a part in making this research effort become a reality. I am enormously appreciative to these individuals who dedicated time and effort to help me and to those who have given me the opportunity to accomplish it.

Firstly, I would like to express my deepest appreciation to my advisor, Dr. Bin Zheng, for supervising and mentoring me throughout my Ph.D. studies. He gave me the opportunity to participate in various research projects and work at great institutes such as OU Norman, OUHSC, Mercy Hospital and Roche. Without his continuous encouragement, guidance and support, I would not be where I am now.

I would like to thank my committee members, Dr. Hong Liu, Dr. Joseph Havlicek, Dr. Tomasz Prezebinda, Dr Liangzhong Xiang and Dr. Han Yuan, for their guidance and suggestions that actuate the progress and improvement of this project.

I would like to thank my supervisor, Dr. Yao Nie, for the mentorship and guidance that she provided me during my internship at Roche.

I would like to thank Dr. Maxine Tan, Dr. Yunzhi Wang, Seyedehnafiseh Mirniaharikandehei, Gopichandh Danala, Morteza Heidari and Linsheng He, my lovely colleagues, for all their collaboration, generous help, and support over the years that lead to the completion of this project. I would like to thank my friend, Dr. Salma Mahzoon for her kindness and friendship.

Finally, I would like to thank my mom and dad, Fateme and Ahmad, my siblings, Farimah, Faride and Amir, for their constant support and unconditional love and my lovely husband, Zheng, for all his supports and understanding.

To my mom, for all her sacrifice for her children.

Table of Contents

List of Tables

List of Figures

Abstract

Reading medical images to detect and diagnose diseases is often difficult and has large inter-reader variability. To address this issue, developing computer-aided detection and diagnosis (CAD) schemes or systems of medical images has attracted broad research interest in the last several decades. Despite great effort and significant progress in previous studies, only limited CAD schemes have been used in clinical practice. Thus, developing new CAD schemes is still a hot research topic in medical imaging informatics field. In this dissertation, I investigate the feasibility of developing several new innovative CAD schemes for different application purposes. First, to predict breast tumor response to neoadjuvant chemotherapy and reduce unnecessary aggressive surgery, I developed two CAD schemes of breast magnetic resonance imaging (MRI) to generate quantitative image markers based on quantitative analysis of global kinetic features. Using the image marker computed from breast MRI acquired pre-chemotherapy, CAD scheme enables to predict radiographic complete response (CR) of breast tumors to neoadjuvant chemotherapy, while using the imaging marker based on the fusion of kinetic and texture features extracted from breast MRI performed after neoadjuvant chemotherapy, CAD scheme can better predict the pathologic complete response (pCR) of the patients. Second, to more accurately predict prognosis of stroke patients, quantifying brain hemorrhage and ventricular cerebrospinal fluid depicting on brain CT images can play an important role. For this purpose, I developed a new interactive CAD tool to segment hemorrhage regions and extract radiological imaging marker to quantitatively determine the severity of aneurysmal subarachnoid hemorrhage at presentation and correlate the

estimation with various homeostatic/metabolic derangements and predict clinical outcome. Third, to improve the efficiency of primary antibody screening processes in new cancer drug development, I developed a CAD scheme to automatically identify the non-negative tissue slides, which indicate reactive antibodies in digital pathology images. Last, to improve operation efficiency and reliability of storing digital pathology image data, I developed a CAD scheme using optical character recognition algorithm to automatically extract metadata from tissue slide label images and reduce manual entry for slide tracking and archiving in the tissue pathology laboratories.

In summary, in these studies, we developed and tested several innovative approaches to identify quantitative imaging markers with high discriminatory power. In all CAD schemes, the graphic user interface-based visual aid tools were also developed and implemented. Study results demonstrated feasibility of applying CAD technology to several new application fields, which has potential to assist radiologists, oncologists and pathologists improving accuracy and consistency in disease diagnosis and prognosis assessment of using medical image.

Chapter 1: Introduction

1.1 Background

Computer-Aided Detection and Diagnosis, known as CAD, involves the application of computerized analysis to assist the radiologist in more accurately and efficiently reading and interpreting different medical images. CAD popularity is growing in the last two decades and is becoming an important area of research in medical imaging informatics field, which aims to assist clinicians (i.e., radiologists and pathologists) in reading and interpreting medical images for disease detection and diagnosis. Traditionally, the clinicians are subjected to do the interpretation process and not always make the optimal use of the data acquired by an imaging device [1, 2]. However, there is a number of limitations for the clinicians to do the image processing and feature or pattern analysis in reading medical images because (1) human eye and brain may not be able to reliably detect and compare many image features among different images in a large scale dataset, and (2) there is large intra- and inter-reader variability in reading and interpreting medical images due to the different training and clinical practice experience of different clinicians, [3-5]. In order to address these issues or challenges, many researchers have been working to develop and apply image processing techniques and CAD systems aiming to assist clinicians (i.e., radiologists and pathologists) to more accurately read and diagnose medical images. CAD systems with the help of medical image processing techniques are able to modify the image presented to the readers in different ways such as enhancing the abnormalities or segmenting a specific lesion and presenting information regard the lesion to assist radiologists in reading medical images. For example, computer-

aided detection (CADe) systems are developed to detect and mark regions of abnormalities to alert radiologists to pay more attention to look and analyze these regions during image interpretation, while computer-aided diagnosis (CADx) aims to assist clinicians in diagnosis of diseases and staging of disease severity or types. Some integrated CAD schemes use CADe to detect and identify the abnormal regions and then used CADx to provide quantitative information and analysis of the detected regions. CAD systems have been widely investigated and developed using variety of image modalities such as conventional digital X-ray images (i.e., mammography), magnetic resonance imaging (MRI), computed tomography (CT), ultrasound and digital pathology (DP) images to detect and diagnose many diseases including cancers such as breast, brain, lung, colon cancer and etc. Different tasks such as segmentation of region of interest and abnormalities (i.e., tumor, hemorrhage, cells), classification of malignant and benign lesions and prediction of treatment response can be conducted using CAD schemes.

For example, in CAD of breast cancer, Dheeba et al proposed a CAD scheme for detecting breast cancer in mammograms using particle swarm optimized wavelet neural network [6], Sun et al, applied a three-class classification approach by using a two-stage classifier combined with support vector machine (SVM) learning algorithm to classify breast tumors detected on mammograms [7], Song et al investigated and compared the efficiency of CAD systems for MRI and other breast imaging modalities in the assessment of tumor extent, lymph node status in patients with invasive breast cancer [8], Yuan et al applied a multimodality CAD scheme that combines image information from full-field digital mammography (FFDM) and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) for breast tumor or lesion classification [9]. These studies

demonstrated that applying CAD systems of different types of medical images using the advanced image processing and machine learning technologies may play an important role to automatically segment, detect, and classify abnormalities to help more accurately and robustly diagnose human diseases using medical images.

1.2 Medical Imaging Modalities

The overall aim of medical imaging can be summarized under several general headings. Human with cancer or other diseases who need immediate treatment requires medical imaging to help clinicians make correct diagnosis and select optimal treatment options. Imaging modalities play an important role in all aspects of disease treatment, including screening, detection, diagnosis, image-guided biopsy, treatment planning, outcome assessment and follow-up examinations [10]. Different imaging modalities are used for different disease, mammography is used as the standard techniques for breast cancer patients, colposcopy images are used for cervical cancer patients, CT images detect the lung cancer and digital pathology images are used by pathologists for detecting tumor cells and diagnosing tissue specimens. In this dissertation, I developed and evaluated CAD schemes of three different image modalities including MRI of breast, CT images of brain and digital pathology images of cell and tissue specimens. The basic imaging characteristics are described in the following subsections.

1.2.1 Magnetic Resonance Imaging (MRI)

Mammography and ultrasonography are the standard imaging techniques for the detection and characterization of breast tumors, while mammography conceives tissue

density differences and ultrasonography demonstrates the difference between solid masses and cysts. Mammography sensitivity, when performed optimally, is between 69 and 90% [11] while the specificity ranges from 10-64% [12]. Observer error, suboptimal technique, and large heterogeneity of the tumor and surrounding fibro-glandular tissue are the cause of low sensitivity and specificity of mammography.

Although mammography is well recognized as the most cost-effective imaging modality for breast cancer screening and early detection, the performance (including both sensitivity and specificity) of screening mammography is not satisfactory [13,14] in particular for younger women with dense breast tissues. As a result, other new imaging modalities have also been developed and tested for improving breast cancer screening and detection. Among them, dynamic contrast-enhanced breast magnetic resonance imaging (DCE-MRI) has shown superior sensitivity in detecting mammography-occult cancers and is recommended by the American Cancer Society as an adjunct screening tool to mammography for women at an elevated breast cancer risk (e.g. >20–25% of lifetime risk) since 2007 [15]. Currently, breast DCE-MRI has been playing an important role in the clinical management of breast cancer in screening, diagnosis and assessment of treatment efficacy [16, 17].

1.2.2 Computed Tomography (CT)

Computed tomography (CT) is a medical imaging procedure that uses ionizing radiation to produce cross-sectional images of the human body. The method by which x-rays are produced within a CT scanner is similar to general X-ray radiography. However, CT scanners have the capability to produce three-dimensional images unlike the two-

dimensional planar X-ray projection images seen in most radiography studies. Computed tomography plays a very important role in diagnoses of many diseases and has been used in a variety of medical imaging procedures because of its unique ability to offer clear images of bone, muscle, blood vessels, and different types of tissue. Where other imaging techniques are much more limited in the types of images they can provide, the role of CT in the clinical setting is extensive. It can be used to plan certain surgeries, guide biopsies, measure bone mineral density, detect injuries to internal organs, and has proven to be a valuable tool for the diagnosis and treatment of many musculoskeletal disorders. CT imaging is even used for the diagnosis and treatment of certain vascular diseases. Probably the most important aspect of CT however, is its role in cancer detection, diagnosis and assess treatment efficacy. It allows physicians to accurately detect and locate different types of cancers and plays an important part in radiation treatment planning process. The importance of CT technology is without a doubt a vital aspect for the diagnosis of diseases and treatment of patients. As the advancements in developing new CT imaging technology and novel imaging procedures (i.e., thin slice, low dose and perfusion imaging), patient care will continue improving.

1.2.3 Digital Pathology (DP)

Pathology provides a ground-truth in disease diagnosis. In order to improve efficacy of disease diagnosis, digital pathology has been attracted broad research interest and developed quickly in medical imaging field. Specifically, instead of visually examining pathological specimens under optical microscopes, digital pathology creates an image-based information environment enabled by computer technology that enables

the acquisition, management and interpretation of pathology information generated from digitized images of tissue specimens placed on the glass slides. Whole slide scanner technology has been increasingly used in the pathology laboratories and a large number of tissue slides are being scanned and processed in recent years. It is well known that useful information is in pathology images and digital pathology has the potential to help improve efficiency of clinical workflow and provide potential of quantitative feature analysis to improve diagnostic accuracy. Additionally, it can also reduce the need for storing large number of glass slides on-site and the risk of the physical slides getting broken or lost. All images and data can be stored digitally in the computer database, which can also be much easy to retrieve by physicians for disease diagnosis and conducting research or educational purpose by medical researchers.

Due to colorless and transparency of a large number of cells, histological sections have to be stained in some ways to be visible. Different staining techniques such as non-specific, staining most of the cell in the same way, or selectively staining of particular chemical groupings of molecules within cells and tissues are available and used in clinical practice. Staining process work by using a dye and counterstain which dye stains some of the cell's components as a bright color and counterstain stains the rest of the cell in a different color. Acidic dyes react with cationic or basic components in cells. Proteins and other components in the cytoplasm are basic and will bind to acidic dyes.

The most commonly used stain in histology is called Hematoxylin and Eosin (H&E). H&E contains two dyes of hematoxylin and eosin which eosin act as an acidic dye, which is negatively charged and stains the basic structures into red or pink.

Hematoxylin act as a basic dye and is used to stain acidic structure into a purplish blue. Thus, the nucleus is stained purple in the H&E staining process. Figure 1 demonstrated an example of H&E staining.

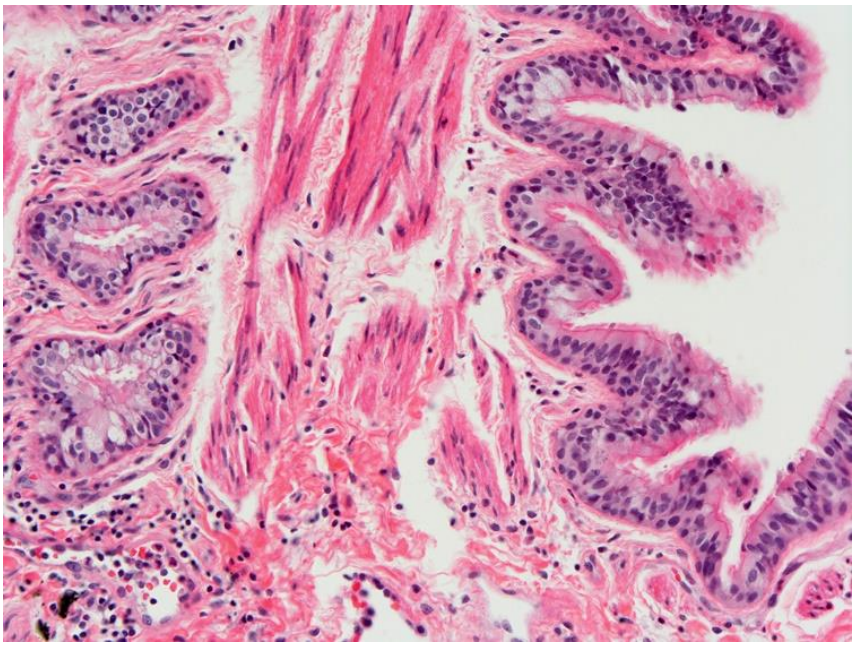


Figure 1. H&E staining example.

1.3. Machine Learning

Machine learning is a field that focuses on the construction of algorithms to make predictions based on training data. A machine learning task aims to identify (to learn) a function *f* that maps the input domain X (of data) onto output domain Y (of possible predictions) [18]. Functions *f* are chosen from different function classes, dependent on the type of learning algorithm that is being used. Mitchell defines "learning" as follows: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" [19]. The performance measure P tells the researchers

quantitatively how well a certain machine learning algorithm is performing. For a classification task, the accuracy of the system is usually chosen as the performance measure, where accuracy is defined as the proportion for which the system correctly produces the output. Experience E that machine learning algorithms undergo are datasets. These datasets contain a set of examples that are used to train and test these algorithms.

Machine learning can be accomplished using either a supervised or an unsupervised approach. In supervised learning, the system receives a dataset with different example parameter values and decisions/classification, from which it infers a mathematical function, which automatically maps an input signal to an output signal. Thus, it figures out what it is supposed to do. The supervised learning method was used in all the studies mentioned in this dissertation.

Unsupervised learning, on the other hand, means that the system acts and observes the consequences of its actions, without referring to any predefined type cases other than those previously observed. This is pure 'learning by doing' or trial-and-error. Compared to supervised learning, unsupervised methods typically perform poorly in the beginning, when they are untuned, but as they tune themselves, performance increases. It can be argued that using unsupervised learning, a classifying system should be able to set up hypotheses that no human can figure out, due to their complexity. If unsupervised methods were used for the studies in this dissertation, the machine learning systems would have to find out the learner stage hypothesis all on its own, which would probably require much more training data than is available.

To evaluate classifier performance given by a machine learning scheme, either a special testing dataset or a cross validation technique may be employed. A test dataset

contains pre-classified data different to those in the training dataset, and is used only for evaluation, not for training. If data are scarce, it is sensible to use cross-validation in order not to waste any data, which could be useful to enhance classifier performance; all data are used both for training the classifier and for testing its performance.

More data do not necessarily mean better classifier performance. Even though the classifier becomes better on the training set it could actually perform worse on the testing data. This is due to the overfitting of the classifier transfer function, so that it fits too tightly to the training data and the border between classes is jagged rather than smooth, unlike how it usually should be.

Although a large number of machine learning classifiers have been developed and are currently available, my studies mainly selected and used following machine learning algorithms or classifiers, which are Support Vector Machines, k Nearest Neighbor, Artificial Neural Network and Decision Tree. The basic learning concept and characteristics of these machine learning classifiers are described briefly as below:

1.3.1 Support Vector Machines

Support Vector Machines (SVMs) are one of the most popular classification algorithms to perform supervised classification over the dataset [20]. The aim of SVM is to find the most optimal classification function that differentiates between units of classes in training data. With a linearly separable dataset, the most optimal classification function can be decided by constructing a hyperplane which maximizes the margin between two datasets and thus creates the largest possible distance between datasets. Figure 2 demonstrates the visualization of SVM strategy.

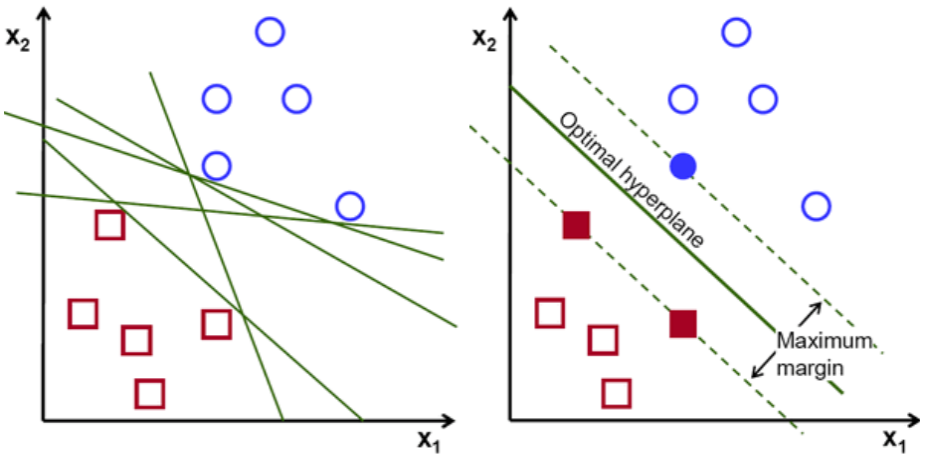


Figure 2. Creating the optimal separating hyperplane with the use of support vectors to classify the two classes.

The idea behind SVMs is that by finding the maximum margin, which represents the most optimal hyperplane and has the best generalization ability that it can reach. This results in the best classification performance for both the training data as well as future data. Thus, a Support Vector Machine is strictly not a machine, but a simple and powerful algorithm.

1.3.2 K Nearest Neighbor

k-Nearest-Neighbor classification, or kNN, is another popular machine learning algorithm that localizes a group of k objects in a training case that has the closest proximity or similarity to the test object, and then assigns a label derived from the prevalence of a class in the closest proximity. Three important components are needed for this algorithm: a group of labeled objects; a proximity metric; and the number k of nearest neighbors [21]. A popular proximity metric that is used for kNN classification is "Euclidian Distance" [22]. Other metrics exist for defining the distance between instances

of a dataset. Examples include the Minkowsky, Camberra or Chebychev metrics [23] although often weighing strategies are used that alter the voting influence for more accurate results.

1.3.3 Artificial Neural Network

Neural networks have gained widespread recognition as an effective machine learning algorithm by outperforming many algorithms such as Support Vector Machines in various relevant applications such as pattern recognition [20, 24]. A neural network is an architecture that comprises of units named neurons. These architectures usually consist of three different layers: the input layer which contains the input feature vector; the output layer that consists of the neural network response; and the hidden layer that contains the neurons that connect to both the input and output. An example of a neural network is illustrated in Figure 3. This artificial neural network, called a Feed-forward neural network, only allows signals to travel from input to output.

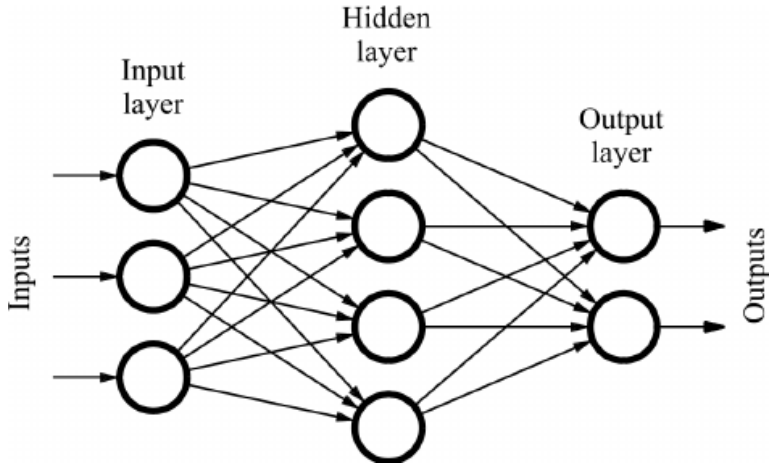


Figure 3. An example of Neural Network design consisting of 3 layers of Input, Hidden and Output Layer.

Artificial neural networks consist of three fundamental characteristics: the network architecture; input and activation functions; and the weight of input connections. The network architecture and functions are chosen at the initial stage and remain the same during training. The performance of the neural network is reliant on the value of the weights. The weights are tuned during training so that a certain output is achieved. ANN can be trained using a multitude of different training programs [25]. A very prominent training method is the back-propagation algorithm [26]. Other techniques include the weight-elimination algorithm that automatically infers the network topology, and genetic algorithm that tries to derive the network architecture and train its weights through competition and evolution of GA-chromosomes through multiple generations [27- 29].

1.3.4 Decision Tree

The Decision tree algorithm comprises of trees that categorize data by looking at feature values. Every node in a decision tree depicts a feature that has to be classified, and the tree branches depict values that are considered by such a node. A survey of existing work on decision tree construction is done by Murthy [30], who describes the use of decision trees in multiple disciplines such as statistics, pattern recognition and machine learning. One of the most popular decision tree algorithms in literature is the C4.5 algorithms by Quinlan [31], an improved version of the earlier decision tree algorithm by Quinlan, the ID3 [32.]. C4.5 uses a divide and conquer strategy to grow a tree, selecting one feature with a minimum of two outcomes that divides the set of samples most effectively. When all instances in a certain set belong to one class, or the set size is too small, a label is assigned to a decision tree leaf equal to the most popular class label [21]. An example of a decision tree is illustrated in Figure 4.

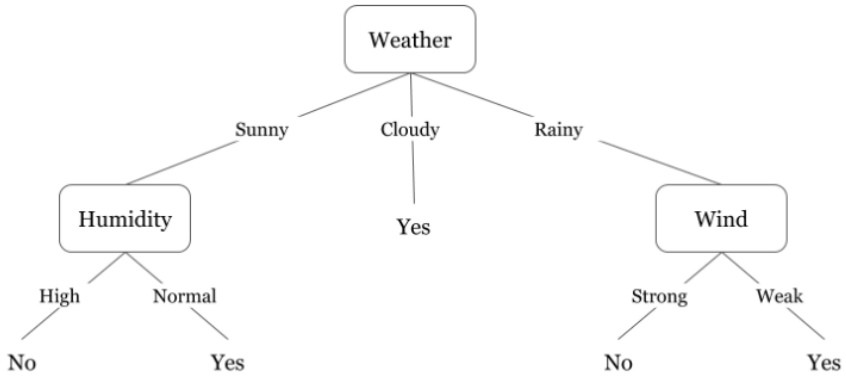


Figure 4. Example decision tree.

The highest decision node is called the root node. The feature that can sort the data most effectively is chosen as the root node. This strategy is duplicated for each sub-division of the training data, until all data is divided into specific class batches. Determining which feature sorts the training data most effectively can be done through many techniques such as the "Gini Index" [33] "Information Gain" [34], or the "ReliefF Algorithm" [35]. Even though each of these techniques differs from the other, a survey of decision tree algorithms by Murthy showed that no single best method exists for determining the best sorting feature [30]. However, for a specific dataset, comparing individual techniques may provide useful insights for the choice of sorting feature.

In order to build right sized trees, Breiman et al. [33] proposed "pruning", a technique that is used to reduce the complexity of decision trees by removing sections of a tree that provide little information for the classification of instances. To prevent overfitting training data, two common approaches can be used: (1) stopping the training algorithm before a perfect fit is reached; and (2) pruning the decision tree. A survey of

popular pruning methods is presented by Elomaa, who concludes that no single best pruning method exists [36].

1.4 Active Learning

Semi-supervised learning is another machine learning class beside supervised and unsupervised learning. As mentioned earlier, in supervised learning, we learn a mapping from observation to predicted output from a set of (input, output) pairs: each training input has an associated output. In unsupervised learning, there are no known outputs, and the goal is to discover structure in the set of (input) points. To take advantages of both learning approaches, the semi-supervised learning includes a set of (input, output) pairs with the labeled truth and another separate set of unlabeled (input) data. The goal is then to use the structure of the unlabeled data points to learn a better mapping than what we would get from labeled data alone. In a supervised machine learning task, we need to obtain labeled data to train a model. In most medical imaging related studies, obtaining the large image dataset with verified label (ground truth) is quite difficult and expensive in terms of time or labor of the clinicians. Thus, the researchers may choose to label a subset of the available data, but the choice of the subset will affect the quality and performance of the final machine learning model or classifier. The question, then, is how to optimally select the subset of data that will achieve the best model performance. The importance of active learning emerges in applications that deal with large amounts of data; since labeling such data can be very costly and exhausting. Active learning is an iterative machine learning algorithm in which the main issue is to evaluate the informativeness of an unlabeled instance. The active learner is a classifier that is initially

trained on a few labeled instances. Then, iteratively, by its knowledge derived from the labeled data it requests a label for one of the instances of the unlabeled pool of data. Successful active learning should lead to a significant reduction of the amount of training data needed for a supervised learning task with no significant reduction of the machine's performance over a machine trained with a fully-labeled dataset. We indicate most of supervised machine learning algorithms as passive learning in terms of fully-labeled training data. The efficiency of active learning has been qualified by two approaches. The more common approach is the reduction of training data required to reach a certain level of performance. The other approach is to increase the performance of the model for a certain amount of training data [37].

The applicability of either of these approaches depends on the established image datasets. Active learning has been studied in many different contexts [38] that provide a comprehensive literature review on the subject. For example, application of speech recognition [39] employed active learning to select a limited subset of utterances to transcribe from a large amount of un-transcribed utterances. It reduced the number of utterances needed for transcribing by 60% to achieve the same recognition accuracy obtained using random sampling. Another application of active learning is in message filtering. For instance, in spam filtering applications, trustworthy labels for messages can be costly to acquire. Sculley reports that online active learning can significantly reduce labeling and training costs while maintaining high levels of filtering performance [40]. Active Learning has been also employed for the task of classification for documents or media like images, audio and videos. For the classification of media not only the

annotation can be very time consuming and tiring can also be mentally challenging (i.e., sexual child abuse). Figure 5 illustrates the active learning cycle in this problem domain.

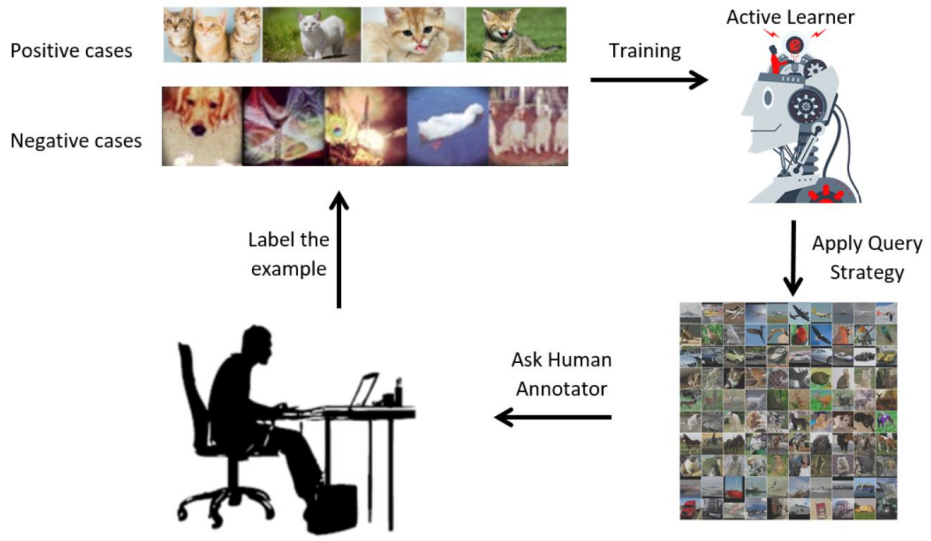


Figure 5. Active learning cycle for image classification task. The cycle starts by labeling few examples as initial seeds then active learner model is created. By applying query strategy on the unlabeled pool of data, the active learner asks the human annotator for the label of the most informative example. The new labeled example will then be added to the training set.

1.5 Objective of Studies

As previously mentioned, CAD systems of medical images have been developed and tested for different tasks such as segmentation and detection of abnormalities, diagnosis of diseases, prediction of prognosis and assessment of treatment efficacy to help provide physicians new quantitative image feature analysis or decision-making supporting tools in current and future clinical practice. Despite great research effort, many technical and clinical challenges remain in developing CAD schemes and expanding CAD application in many different clinical application fields. The goal of this dissertation is to investigate several new and novel CAD concepts and approaches, as well as to develop and test

feasibility of applying these CAD approaches to several new clinical applications using breast MRI, brain CT and digital pathology images. The study hypothesis is that by developing these new CAD schemes, we can identify new quantitative imaging markers, provide clinicians new visual-aid or decision-making supporting tools to better read and interpret medical images, and improve efficacy of using digital pathology technology in the clinical practice. Developing and implementing these new CAD schemes or tools can have potential to help improve accuracy and consistency of the clinicians in disease detection, diagnosis, and patient management and treatment.

1.6 Organization of Dissertation

To test the study hypothesis described in previous subsection, I have conducted many studies in the last several years in University of Oklahoma and Roche Company. In this dissertation, I report four applications of developing CAD schemes using three medical imaging modalities (MRI, CT and Digital Pathology images) to segment disease regions (ROIs) and apply different machine learning algorithms to perform the classification tasks (Chapter. 2,3,4,5) and an application of developing optical character recognition scheme to help convert analog information to digital information and realizing storage of digital cell or tissue specimen images in pathology laboratory (Chapter 6). Specifically, in Chapter 2, a new quantitative global breast MRI feature analysis scheme was developed to predict tumor response to the neoadjuvant chemotherapy using breast MRI images acquired pre-therapy. In Chapter 3, a new quantitative image processing and feature analysis method was developed and evaluated to predict the pathologic complete response of breast cancer patients who underwent neoadjuvant chemotherapy using images acquired post-therapy. In Chapter 4, a

computer-aided detection tool was developed to segment and quantify intracranial radiologic markers using brain CT images. In Chapter 5, a computer-aided detection scheme was developed to automatically identify the non-negative digital pathology slides for IHC assay development. In Chapter 6, a new CAD scheme with optical character recognition algorithm and graphical user interface named Tissue Slide Reader was developed and tested. Last, in Chapter 7, a summary of these new CAD schemes and their future application potentials are discussed, which generate the conclusion of this dissertation including the future work of this candidate.

Chapter 2: Developing a New Quantitative Global Breast MRI Feature Analysis Scheme to Assess Tumor Response to Chemotherapy

2.1 Introduction

Among the existing imaging modalities of breast cancer detection, diagnosis, and prognosis assessment, dynamic contrast enhancement (DCE) breast magnetic resonance imaging (MRI) has superior capability and performance. For example, one review article analyzed five prospective clinical studies and revealed a comparison result in which the cancer detection sensitivities were 40% and 81% for mammography and breast MRI, respectively [41]. Another large clinical study involving 2809 women with elevated breast cancer risk also reported that using breast MRI enabled detection of 79% (41/52) mammography-occult cancers [42]. Hence, using breast MRI could detect a significantly greater number of cancers than mammography. As a result, breast DCE-MRI has been recommended by the American Cancer Society as an adjunct screening tool to mammography for women with lifetime breast cancer risk greater than 20–25% since 2007 [15]. In addition, breast MRI has played an important role in classifying breast tumors (ie, ductal carcinoma in situ [DCIS] or invasive ductal carcinoma [IDC] with and without metastasis positive lymph nodes [43]) and evaluate the tumor response to chemotherapy by comparing two sets of breast MR images acquired pre- and post-chemotherapy [17].

In current practice, advanced stage breast cancer patients with large breast lesions are treated with neoadjuvant (preoperative) chemotherapy before surgery. Based on the tumor response to the neoadjuvant chemotherapy, a breast cancer patient may receive

breast-conserving surgery instead of mastectomy [44-46] or avoid the surgery due to the pathological complete response (pCR) [47, 48], which would improve the life quality of breast cancer patients. Tumor response to the chemotherapy is typically evaluated by comparing tumor size and kinetic feature variation using breast MRI examinations taken pre- and post-chemotherapy based on the RECIST guideline [49]. However, due to the large heterogeneity of breast tumors, the response to the neoadjuvant chemotherapy varies widely in different patients [50]. Hence, in order to assist clinicians in making an optimal treatment plan for the individual patients early and/or reduce the toxicity or other side effects of the patients who do not receive significant benefit from the neoadjuvant chemotherapy, developing a new clinical marker that allows more accurate prediction or assessment of tumor response early may have high clinical impact. For that purpose, we developed and tested a new quantitative kinetic image feature analysis-based CAD scheme using breast MR images acquired before the patients participated in the neoadjuvant chemotherapy to predict tumor response to the neoadjuvant chemotherapy [51]. From the segmented breast lesions, the scheme computed kinetic image features of the lesions and assessed the likelihood of tumor response to the neoadjuvant chemotherapy using a multi feature-based artificial neural network.

Despite encouraging results in our preliminary study, application of the scheme can be limited or less robust due to the difficulty in accurately defining and/or segmenting the breast lesions, in particular the subtle and/or diffuse lesions without a solid lesion boundary. Recently, studies reported by several groups [52,53] demonstrated that the global background parenchyma enhancement (BPE) features extracted from the entire breast MR images can be used to assess breast cancer risk [52] and improve accuracy in

classifying between malignant and benign breast lesions [53]. Hence, based on previously published works, the objective of this study was to develop a new quantitatively global kinetic image feature analysis-based CAD scheme without lesion segmentation and test the feasibility of applying this new scheme to assess tumor response to the neoadjuvant chemotherapy.

2.2 Materials and Methods

2.2.1 DCE-MRI Image Dataset

The study protocol including an image data collection and data analysis method was approved by our Institutional Review Board. In this retrospective study, the image data were collected from the existing cases stored in the clinical Picture Archiving and Communication System (PACS). The informed consent of the patients was waived. The dataset includes the deidentified breast MR images acquired from 151 breast cancer patients. Each patient had two sets of breast DCE-MRI examinations taken before and after neoadjuvant chemotherapy. The average time difference between two MRI examinations is around 5 months (or 157 days). All MRI examinations were performed using a 1.5T GE Excite MRI scanner in 2008 and 2010. In each MRI examination, five sets of axial images were scanned and acquired. The first one is a pre-contrast scan of two breasts. Approximately 3 minutes after intravenous administration of 0.1 mmol/kg body weight OptiMARK gadolinium, four sets of post-contrast scans started to acquire four sets of new axial images. Each image slice has 512 × 512 pixels with a pixel size of 0.6445 × 0.6445 mm and a slice thickness of 4 mm.

Among the 151 patients, 63 were assigned to the complete response (CR) group and 88 were categorized as the partial response (PR) group. According to RECIST guidelines, in the CR group the kinetic enhancement signals (or contrast-enhanced pixels) inside the tracked breast tumors depicting the first set of breast MR images acquired pre-chemotherapy become undetectable in the second set of breast MR images acquired post-chemotherapy, while in the PR group the contrast-enhanced pixels inside the tumor regions decrease by more than 30% between the post- and pre-chemotherapy breast MR images. Figure 6 shows three examples of two matched breast MR image slides acquired pre- and post-chemotherapy of three patients. In one CR case (Fig. 6 a, b), the contrast-enhanced pixels almost disappear inside the target tumor region depicted on the post-chemotherapy breast MR image. In the first PR case (Fig. 6 c, d) a solid tumor has contrast enhanced pixels surrounding a tumor boundary and a big necrotic center region. After neoadjuvant chemotherapy, although the tumor diameter was reduced or "partially responsive," the central necrotic region of the tumor also disappears and the active contrast enhancement volume increases after the neoadjuvant chemotherapy. The second PR case (Fig. 6 e, f) shows the lesions with diffused enhancement on breast MR images acquired pre- and post-chemotherapy.

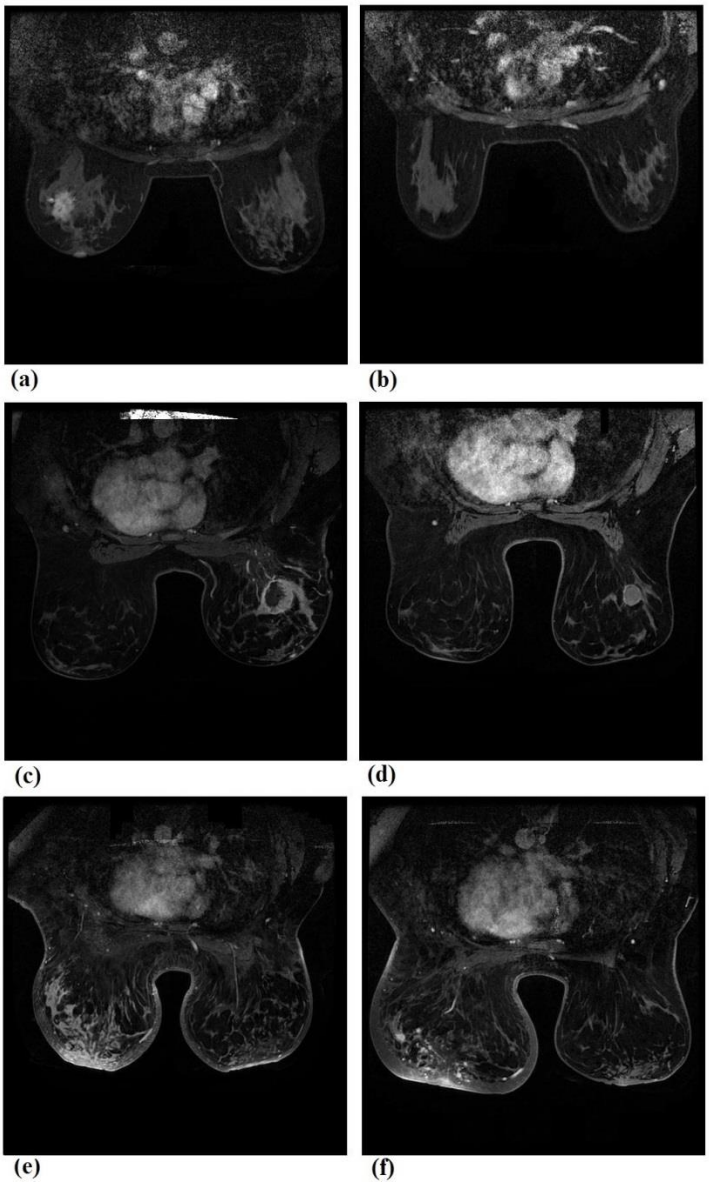


Figure 6: Three examples of showing two matched breast MR image slides acquired from pre- and post-chemotherapy of one CR case (a) and (b), one PR case with solid contrast enhanced tumor (c) and (d), and one PR case with the diffused lesion enhancement (e) and (f).

In this dataset, the age of the patients ranged from 25 to 76 years old. The average age and standard deviation are 47.4±11.2 and 49.2±10.4 for CR and PR patient groups,

respectively, which indicate that the majority of women whose breast DCEMRI examination images were selected in this study are relatively younger (<50 years old).

Table 1 summarizes the basic tumor characteristics between the CR and PR case groups. The majority of tumors were diagnosed as invasive ductal carcinoma (IDC) with or without associated DCIS. Finally, the axial view of the breast MR images acquired pre-chemotherapy of each patient was used and analyzed in this study. The goal of this study was to use the global kinetic image features computed from the pre-chemotherapy breast MR images only to build a machine-learning classifier or model to assess tumor response to the neoadjuvant chemotherapy.

Table 1: Distribution of Tumor Subtypes in Both CR

Tumor characteristics	**CR**	**PR**
Number of solid tumors	25	43
Average size of solid tumors		
Number of diffused tumors	38	45
Tumor pathology subtypes		
Invasive ductal carcinoma (IDC)	26	27
IDC and DCIS	28	45
Invasive lobular carcinoma (ILC)	2	1
ILC and DCIS	2	1
Infiltrating carcinoma	0	1
Invasive mammary carcinoma	0	3
Poorly differentiated carcinoma	5	10

2.2.2 Quantitative Global Kinetic Image Feature Computation

In order to automatically compute global contrast enhancement features, we applied a CAD scheme of breast MR images that has been developed and to segment the breast region depicted on each breast MR image by removing all pixels in the air background and behind the chest wall or pectoralis muscle. The details of this CAD scheme have been reported in our previous publication [51]. In brief, the CAD scheme uses the following four image processing and feature computation steps: 1) applying a threshold method to remove all pixels in the air background of each image; 2) detecting the chest wall depicted on each breast MR image to remove all pixels behind the chest wall (Figure. 7); 3) performing image registration and subtraction of two sets of matched breast MR image slices acquired in two breast MRI scans performed pre- and post-injection of gadopentate dimeglumine (Gd-DTPA) contrast agent; and 4) computing a set of relevant kinetic image features from segmented breast regions depicted on each subtraction image. By applying this CAD scheme to all breast MR image slides, we are able to yield the global breast contrast enhancement image features of the whole breast volume.

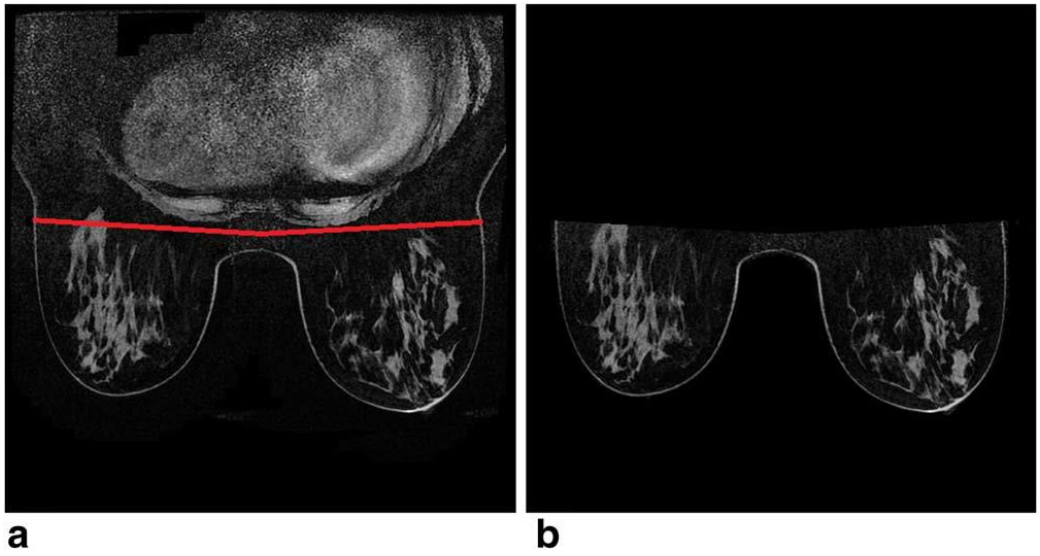


Figure 7: An example to show chest wall detection and breast region segmentation, in which (a) shows an original image slice marked by two CAD scheme detected lines to segment between two breasts and chest regions, and (b) shows the final segmented breast regions, which are used to compute the BPE features.

Our CAD scheme initially computed a set of 10 kinetic or contrast enhancement image features from the subtraction images of two sets of registered images acquired pre- and post-injection of contrast agent (as shown in Table 1). Specifically, these features include: F1: the average contrast enhancement value (EV), which is simply computed as an average of the pixel value of all pixels inside the segmented breast volume; F2: standard deviation and F3: skewness of the contrast enhancement values, which are two features that measure the heterogeneity of contrast enhancement of the pixel values; F4: the maximum contrast enhancement value inside the whole breast volume. In addition, CAD scheme sorted the contrast enhancement values from the maximum to the minimum, which are computed from the whole breast volume, and then computed two new features: F5: the average contrast enhancement value among the pixels listed in the top 1%, and F6: in the top 5% of the sorting list. In a previous study, using these two average contrast

enhancement values could have higher discriminatory power to assist classifying between malignant and benign breast lesions than using the average or the maximum contrast enhancement value computed from the whole breast regions [54]. In addition, the CAD scheme also computed four features representing the bilateral differences (or the absolute subtraction) of two feature values computed from the left and right breast regions (F7–F10).

2.2.3 Feature Analysis and Performance Assessment

Next, from the initial feature pool of 10 contrast enhancement image features (Table 2), we aimed to select optimal image features and build a machine-learning classifier to assess the likelihood of the breast tumors being completely responsive to the neoadjuvant chemotherapy (CR). For this purpose, we used a publicly available data mining and machine-learning software platform, Weka [55], to perform feature selection and classifier training and testing tasks. A similar optimization process has been applied and tested in our previous study [51]. In brief, we used a specific machine-learning classifier, the "Attribute Selected Classifier," which integrates an artificial neural network (ANN) as the base classifier and a Wrapper Subset Evaluator (WSE) to guide feature selection from the initial pool of 10 features. This integration takes a search algorithm and evaluator next to the base classifier, which makes the feature selection process transparent and the base classifier operates only in a reduced optimal feature space [56].

Table 2: Description of 10 computed global kinetic image features.

Feature	Description	Feature	Description
F1	Average enhancement value (EV)	**F6**	Average EV of top 5%
F2	Standard deviation of EV	**F7**	Bilateral average EV difference
F3	Skewness of EV	**F8**	Bilateral STD EV difference
F4	Maximum EV	**F9**	Bilateral difference of average EV of top 1%
F5	Average EV of top 1%	**F10**	Bilateral difference of average EV of top 5%

Due to the limited dataset size of 151 cases, the classifier was trained using a leave-one-case-out (LOCO) method to maximally use all available training cases and also minimize the testing bias [57]. In addition, to avoid bias in feature selection, the feature selection process was embedded inside the LOCO training and testing iteration loops. In each LOCO training and testing process, one case was selected as an independent testing case and the remaining 150 cases were used as training cases. The WSE guided feature selection method was applied to all 150 training cases to search for optimal features from the entire feature pool and train an ANN-based classifier. The trained ANN was then applied to one independent testing case and generated a classification score for the testing case (ranging from 0 to 1). The higher score indicates a higher probability of the tumor responding to the neoadjuvant chemotherapy or being classified into the “CR” class. This

LOCO process was repeated 151 times. As a result, 151 classification scores were independently generated for all 151 cases in our dataset. In the different LOCO training and testing iteration cycles, the potentially different image features may also be selected from the initial pool of features and used to build the ANN.

We then used the area under a receiver operating characteristic (ROC) curve (AUC), which was computed using a publicly available maximum likelihood-based ROC curve fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/, University of Chicago), as an evaluation index to assess performance of the image features or ANN-generated classification scores associated with the tumor response to the neoadjuvant chemotherapy. The *P*-values were also computed by the ROCKIT program when any two sets of image feature data and/or ANN classification scores were used as input data of the ROCKIT program. All statistically significant differences were defined as $P < 0.05$. We compared the discriminatory power of using each individual features (as listed in Table 1) and the classification scores generated by the multi feature based machine-learning classifier (ANNs) using the LOCO validation method. In addition, we applied an operation threshold of $T = 0.5$, which is the middle point of the ANN-generated classification scores to generate a confusion matrix from which we computed the overall classification accuracy as well as the predictive values of both the "CR" and "PR" classes of our scheme in assessing the tumor response to the neoadjuvant chemotherapy.

2.3 Results

When applying each of 10 individual image features computed in our initial feature pool (Table 2) to associate with or classify cases between the CR and PR groups, Table 3 shows and compares the computed AUC values. The AUC values ranged from 0.542±0.047 to 0.734±0.043. Among them, the average contrast enhancement value of entire breast volume or regions (F1) and standard deviation (F2) have the highest discriminatory power, with AUC >0.7.

Table 3: AUC values of applying 10 individual features (given in Table 1) to classify between CR and PR group of cases.

Feature	AUC	P-value	95% CI
F1	0.734±0.043	**F6**	0.623±0.046
F2	0.723±0.043	**F7**	0.572±0.047
F3	0.695±0.508	**F8**	0.613±0.046
F4	0.581±0.047	**F9**	0.542±0.047
F5	0.596±0.047	**F10**	0.542±0.048

Table 4 summarizes the percentage of each of 10 features selected during the 151 LOCO training and testing iteration cycles in building the ANN-based classifier. Despite high AUC values of F2, it was not selected much during LOCO training and testing iteration cycles because of high correlation with other features (e.g., its correlation coefficient with F1 is 0.984). The top four selected features that were selected 90% times in 151 LOCO iteration cycles were the average enhancement value of the entire breast area (F1), average of bilateral enhancement value difference between left and right breasts

(F7), standard deviation of bilateral enhancement value difference between left and right breast (F8), and average of bilateral enhancement value of the top 5% difference between left and right breast (F10). This indicates that these four features played the most important interaction role in developing our ANN based classification model. Three of these features were computed from the bilateral asymmetry of the contrast enhancement features computed between the left and right breasts. Although using each individual image feature may only have limited discriminatory power (or AUC values), developing a multiple feature fusion-based ANN classifier enabled yielding a significantly higher assessment performance with AUC = 0.83±0.04 than using each feature individually ($P < 0.05$). Figure 8 shows and compares five ROC curves generated using the classification scores generated by the ANN-based classifier and each of four commonly selected image features.

Table 4: A list of 9 image features that were selected in LOCO training and testing iteration cycles to test 151 testing cases in our dataset.

Feature	Percentage	Feature	Percentage
F1	100%(151/151)	**F6**	60%(91/151)
F2	36%(55/151)	**F7**	100%(151/151)
F3	0%(0/151)	**F8**	100%(151/151)
F4	11%(16/151)	**F9**	56%(85/151)
F5	62%(93/151)	**F10**	90%(139/151)

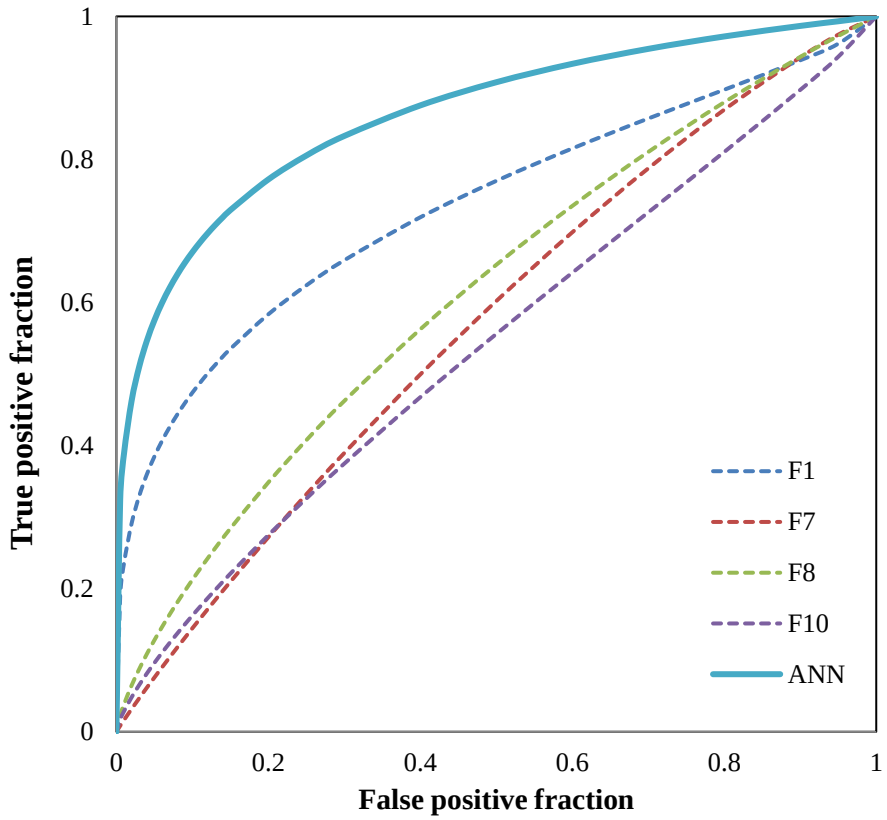


Figure 8: Comparison of 5 ROC curves generated using the classification scores of the ANN-based classifier (solid curve) and other four individual features (dashed curve).

As an example, Table 5 presents the computed feature values and ANN-generated classification scores of three cases (as shown in Fig. 6). The CR case has the highest classification score. Except feature F10 of case 3, the trend between ANN-generated classification scores and other feature values are also demonstrated. Table 6 shows a confusion matrix that was obtained by applying an operation threshold of 0.5 to the classification scores generated by the ANN-based classifier. The overall assessment accuracy of these two groups of 151 CR and PR cases was 82%, in which 124 cases were correctly classified into "CR" and "PR" classes, while the remaining 27 cases were incorrectly classified. The predictive values of "CR" and "PR" case groups are 86.0% (43/50) and 80.2% (81/101), respectively.

Table 5: Feature Values of Three Cases with Images Shown in Figure 6.

Feature	Case 1 (up)	Case 2 (middle)	Case 3 (bottom)
F1	0.29	0.39	0.6
F7	0.32	0.66	1
F8	0.37	0.6	0.76
F10	0.48	0.83	0.42
ANN	0.996	0.254	0.045

Table 6: A confusion matrix of prediction scores generated using an ANN-based classifier that was trained using 4 selected image features.

Actual Cases \ Prediction Result	CR	PR
CR	43	20
PR	7	81

2.4 Discussion

Since breast MRI is the most popular imaging modality used in current clinical practice to assess breast tumor response to neoadjuvant chemotherapy [47, 48], developing and applying breast MRI image feature analysis based on the quantitative image feature analysis or CAD schemes to assess complete response of the breast tumors to the neoadjuvant chemotherapy has been attracting research interest [58, 59]. In this Chapter, we report our latest progress to develop and test new quantitative image feature analysis schemes to predict breast tumor response to neoadjuvant chemotherapy. Our study has a number of unique characteristics. First, we tested the feasibility of whether, with the help of automated segmentation of breast regions depicted on the breast MR

images acquired before pre-chemotherapy, applying a new CAD scheme and machine-learning classifier optimized using the quantitative global kinetic breast MR image features had potential to generate a useful clinical image marker in assessing tumor response to neoadjuvant chemotherapy. Our study results support the hypothesis that the global background parenchymal enhancement (BPE) of breast MRI carries useful clinical information, which may be used to assess breast cancer risk [52] and classify between malignant and benign breast MRI examinations [53]. In this study we demonstrated that using BPE-related image features could also effectively assess tumor response to neoadjuvant chemotherapy.

Second, unlike many previously developed CAD schemes of breast MR images, including our own scheme [51], which computed and used the kinetic image features only from the segmented tumor regions, the CAD scheme developed in this study does not segment the targeted tumors from the breast MR images. As a result, the new CAD scheme is much simpler and probably also more robust to be applied to the different breast MR images with diverse clinical patterns because the automated tumor segmentation process is not only often complicated, but also difficult and/or not robust, in particular for segmenting the diffusive tumors depicted on breast MR images. Although the maximum contrast enhancement typically occurs inside a malignant breast tumor as measured by feature F4, the study results showed that using F4 did not yield the highest classification performance, as comparing several features computed from the global breast MR images including average contrast enhancement (F1) and standard deviation of the contrast enhancement (F2), which indicates that tumor response to the neoadjuvant chemotherapy

depends more on the overall contrast enhancement patterns generated from both tumor and background parenchymal tissues.

Third, besides computing the global BPE features from the entire breast MR images, we also computed bilateral image features asymmetry between left and right breasts. We observed that, unlike in assessing breast cancer risk [60], the computed kinetic enhancement feature difference between two bilateral breast MR images is relatively bigger and classification performance of using such a single feature is lower than using the global BPE features. However, these features are still very useful to build a highly performing multi feature fusion based classifier. When embedding a WSE inside a LOCO-based cross-validation method to select optimal features and optimize the ANN classifier, three bilateral BPE kinetic features asymmetry were among the four mostly selected image features. The results indicate that the applied machine-learning method was effective, which enables selecting and optimally fuse non-redundant image features and eliminate the redundant (or highly correlated) features (i.e., features F1 and F2).

Fourth, this was a retrospective study. We applied our CAD scheme to a relatively diverse image database randomly collected from the existing clinical database. The dataset is relatively balanced in which there are no statistically significant differences in patients' age ($P = 0.445$) and solid tumor sizes ($P = 0.509$) between the CR and PR patient groups. Hence, our study results support a recently emerged Radiomics concept that hypothesized that quantitative image features enabled phenotyping many useful biological or gene-expression processes of cancer development and prognosis [61], which

provides a simple approach to improve decision-making in cancer treatment at low cost [62]. There are also several limitations to this study. For example, 1) this is a laboratory-based retrospective study using a relatively small data size (with 151 cases). Thus, the robustness of our study results needs to be further tested in future studies with large and diverse image datasets. 2) Due to our small dataset, our CAD scheme was only trained and applied to classify cases into CR and PR groups. 3) To make our CAD scheme simple, the scheme does not involve a nonrigid image registration algorithm in an attempt to register MR image slices acquired pre- and post-contrast enhancement scans. However, due to the lack of ground truth, no accurate nonrigid image registration algorithm is available to date. This issue needs to be further developed and investigated. 4) This is a single-institution study. The guidelines of neoadjuvant chemotherapy for breast cancer patients may vary at different medical institutions. Thus, whether our CAD scheme can be optimally applicable to images acquired from other medical institutions also needs to be tested in future studies.

In conclusion, in this study we developed and tested a new CAD scheme based on the quantitative global kinetic breast MR image feature analysis. From our study we demonstrated the feasibility of identifying a new image feature based clinical marker to assess breast tumor response to the neoadjuvant chemotherapy. Our study results support the new concept of Radiomics [61, 62], in which the high association or supplementary information between the quantitative radiographic image features and genomic biomarkers can be found. As a result, although many genomic biomarkers have been performed in the effort to associate breast tumor response with neoadjuvant

chemotherapy, our study suggests that the quantitative image features computed from breast MR images also enable providing highly discriminatory information, which can be more easily extracted from the existing diagnostic breast MR images and yield high clinical impact and cost-effectiveness.

Chapter 3: Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy Using Quantitative Breast MR Images Feature Analysis

3.1 Introduction

Although neoadjuvant chemotherapy is now increasingly used as the first line treatment in patients diagnosed with locally advanced breast cancer in order to downstage tumor, optimize surgical outcomes and reduce risk of cancer recurrence [63], it also faces several clinical challenges. A recent international survey reported that up to 27% of new breast cancers are currently treated with neoadjuvant chemotherapy [64]. Due to the large heterogeneity of breast tumors, response to neoadjuvant chemotherapy varies significantly among the breast cancer patients. In the clinical practice, up to 30% of patients underwent neoadjuvant chemotherapy may have pathological complete response (pCR) without residual invasive cancer [65]. Studies have also showed that patients with pCR had favorable prognosis and/or quite long term disease-free survival (DFS) [66, 67]. However, due to the lack of accurate prognostic markers, the majority of pCR patients still undergo aggressive surgical interventions, which generates overtreatment and unnecessary side effects to the patients. In order to overcome this clinical challenge of avoiding or minimizing overtreatment, as well as help establish new personalized cancer treatment paradigm, identifying and developing new prognostic clinical markers including the imaging markers has been attracting broad research interesting in recent years [49, 68].

In medical imaging field, dynamic contrast enhanced breast magnetic resonance imaging (DCE-MRI) is a routinely used imaging modality to assess breast tumor response

to neoadjuvant chemotherapy [68, 69]. A recent clinical trial (NRG-BR005) launched within National Clinical Trials Network aiming to explore feasibility of eliminating surgery after successful neoadjuvant chemotherapy indicated the need for accurate and robust prognostic markers to improve clinicians' decision-making as it pertains to surgical and chemotherapy recommendations [70]. Currently, tumor response to the neoadjuvant chemotherapy is typically estimated by the comparing change of the tumor size extracted from dynamic contrast enhanced breast magnetic resonance imaging (DCE-MRI) examinations of pre and post chemotherapy [44] based on the response evaluation criteria in solid tumors (RECIST) guidelines [71]. Accuracy of predicting pCR using DCE-MRI and RECIST criteria still remains lower with approximately 50% negative predictive value [72]. A recent *ACRIN 6657* Trial involving 138 cancer patients treated with neoadjuvant chemotherapy reported that the correlation between longest diameter of tumors measured from MRI and final pathology size of tumors was surprisingly low (r = 0.33). Combining MRI, clinical examination, and mammography yielded the highest pCR prediction performance with the area under ROC curve, $AUC = 0.76$ [73]. Thus, in order to increase accuracy in predicting pCR using the image features or markers extracted from DCE-MRI, more research effort and progress are needed.

In this study, we hypothesized that using a machine learning approach to optimally combine both kinetic and texture based image features computed from DCE-MRI may help yield significantly higher performance or accuracy in predicting pCR of breast cancer patients underwent neoadjuvant chemotherapy. In order to test our hypothesis, we investigated and evaluated a new quantitative image processing and feature analysis method to predict pCR. Specifically, we first used a computer-aided detection (CAD)

scheme to automatically remove chest wall and segment breast region depicting on breast MR images acquired after performing the neoadjuvant chemotherapy and compute a set of kinetic and texture-based image features. Next, a random selection method was used to generate new dataset from the initial dataset. Last, three different machine learning algorithms were trained and tested using ten-fold cross validation method to predict pCR of breast cancer patients underwent neoadjuvant chemotherapy. Four parameters were computed from the result of the prediction to assess the performance of using all features. Figure 9 demonstrates the diagram for this study. The details of our study and experimental results are reported in the following sections.

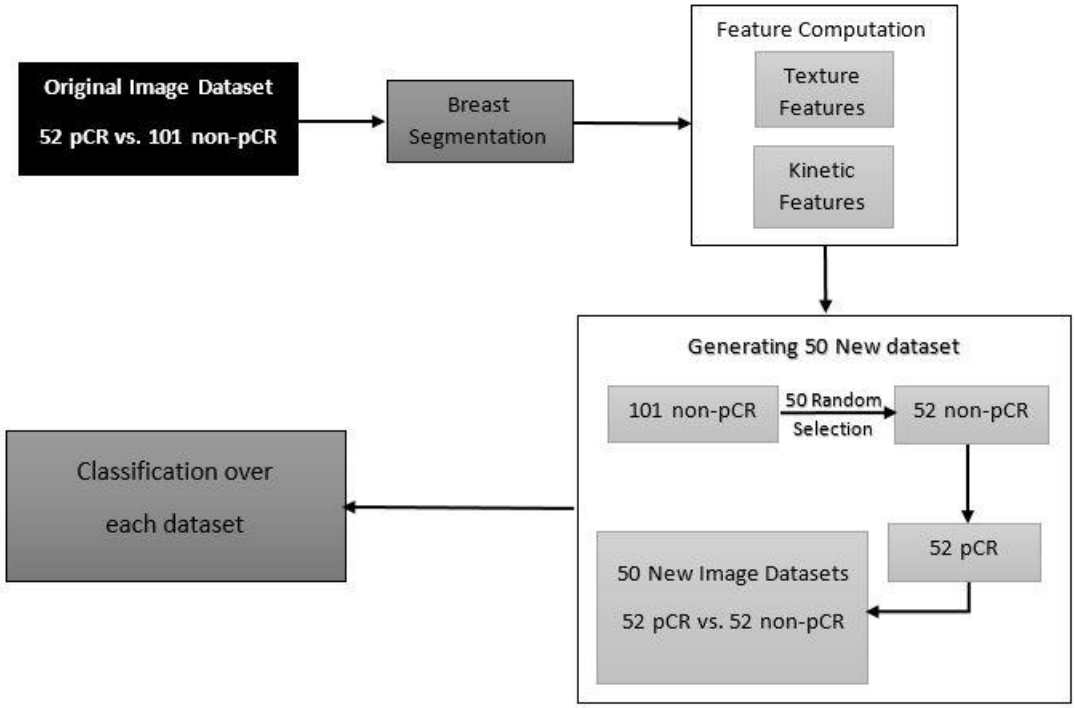


Figure 9: The diagram of subsequent steps in this study.

3. 2 Materials and Methods

3.2.1 Breast MR Image Dataset

A retrospectively collected breast DCE-MRI dataset was collected and used in this study. The dataset includes images acquired 154 breast cancer patients underwent neoadjuvant chemotherapy. Each patient had two subsequent breast DCE-MRI examinations taken before and after the neoadjuvant chemotherapy. The average time difference between two examinations is 157 days. After neoadjuvant chemotherapy, each patient underwent a surgical intervention procedure. Based on the pathologic examination reports, the dataset was then divided into two groups of "pathologic complete response" (pCR) cases and "non-pathologic response" (non-pCR) cases. Among the 154 patients, 53 patients were assigned to the pCR group in which no invasive cancer cells or residual

malignant lesions were detected in the pathological tests, while other 101 cases were assigned to the non-pCR group in which the residual cancer cells were detected and reported in the pathological test reports.

In this dataset, the mean and standard deviations of the patient's age are 50±9.1 and 49±11.1 for the groups of pCR and non-pCR cases, respectively. Each MRI examination in the originally recorded data include five sets of axial view and two sets of sagittal view images acquired before and after the neoadjuvant chemotherapy using a 1.5T GE Excite MRI scanner. The five sets of axial view include one acquired pre-injection of gadopentate dimeglumine (Gd-DTPA) contrast agent and four series of post injection of the contrast agent. From the breast DCE-MRI examination performed after neoadjuvant chemotherapy, the first two MR image scanning series namely, the pre- and the first post-contrast agent injection, were used and analyzed in this study. Each image slice has 512×512 pixels with the pixel size of 0.58 mm in each direction and slice thickness of 2.6 mm.

3.2.2 Image Processing and Feature Computation

In order to develop a new quantitative imaging marker or prediction model to predict pCR cases, this study involves following steps. In brief, a computer-aided detection (CAD) scheme developed and reported in our previous studies [51, 74] was first applied to automatically detect the chest wall and segment breast region depicted on breast MR images. Our CAD scheme include three image processing and feature computation steps namely, (1) segmenting breast region by detection of chest wall and removing all regions behind the chest wall; (2) applying image registration and subtraction method on the two sets of matched breast MR images acquired pre- and the first post-injection of the contrast

agent to generate a new set of contrast enhanced image maps; and (3) computing a set of image features from the generated image maps. Next, a multi-feature fusion based machine learning method was applied to select optimal features and build prediction model. Last, the statistical data analysis method was used to assess the performance of the CAD scheme and pCR prediction model. The details of each step are presented as follows.

A. Breast region segmentation and image registration

Our CAD scheme used a region blobbing algorithm to remove any pixels in the air background of each image. A series of algorithms including Otsu's thresholding method, morphological operation and region growing algorithms were applied to detect the chest wall by generating a separation line and remove any regions (or pixels) behind the chest wall from the breast MR images. The area surrounded between the separation line and breast skin were defined as the segmented breast region and were used to generate a map to be applied on the breast MR images. Next, an image registration algorithm was applied to perform a simple subtraction method on two series of the MRI scans acquired before and after the injection of the contrast agent. A new series of contrast enhancement image maps were generated that include the segmented and registered breast regions. Figure 10 demonstrates an example of the breast region segmentation process and the generated new contrast enhancement image map.

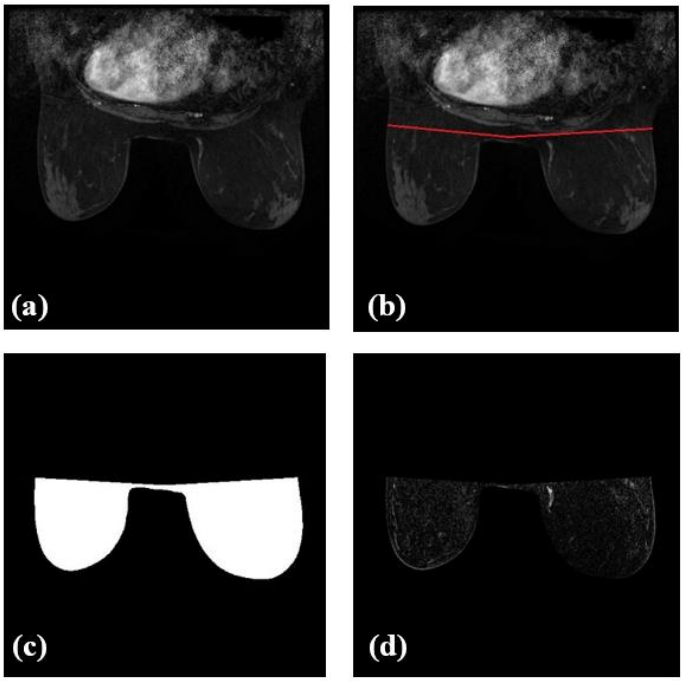


Figure 10: An example of the breast region segmentation steps and generating the contrast-enhanced image map including (a) the original image, (b) separation line, (c) generated mask and (d) the breast region segmented on the contrast-enhanced map.

B. Image feature computation

After image segmentation and subtraction to generate the contrast-enhanced image maps, CAD scheme computed a total of 38 features including both texture and kinetic features from the generated contrast-enhanced images. Among the initial feature pool of 38 features, 28 were texture-related features computed in three-dimensional (3D) space of image maps, which include 17 features derived from gray level co-occurrence matrices (GLCM) and 11 features computed from the gray level run length method (GLRLM). The GLCM method characterizes the texture of an image by calculating how often pairs of pixel with specific values in a specified spatial relationship occur in an image, while the GLRLM method measures the size of homogeneous runs for each gray level in the image.

From the GLCM matrix, CAD scheme computed following texture features, which include energy, contrast, homogeneity and correlation as previously defined by Haralick et al [75] as:

$$\text{Energy} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} g_{ij}^2 \quad (1)$$

$$\text{Contrast} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 \, g_{ij} \quad (2)$$

$$\text{Homogeneity} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} \, g_{ij} \quad (3)$$

$$\text{Correlation} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij) g - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (4)$$

where g_{ij} is the ith and *j*th entry of the GLCM representing a gray-tone intensity in the matrices and μ_x , μ_y and σ_x , σ_y are the mean and standard deviation of the probability matric, respectively.

In these features, energy illustrates the textural uniformity of the image, contrast measures the difference between the lowest and highest voxels gray tones, homogeneity computes the diagonal elements within the displacement vector of the GLCM, and correlation describes the linear dependency on the neighboring gray tone intensities. The remaining GLCM texture features included entropy, autocorrelation, cluster prominence, cluster shade, cluster tendency, dissimilarity, inverse variance, difference entropy, maximum probability, sum average, sum entropy, and variance.

CAD scheme also extracts and computes 11 texture features from the Gray Level Run Length matrices (GLRLM) of the segmented breast region in each slice [76]. The

computed GLRLM features include (1) Short Run Emphasis (SRE), (2) Long Run Emphasis (LRE), (3) Gray Level non-Uniformity (GLN), (4) Run length non-uniformity (RLN), (5) Run Percentage (RP), (6) Low Gray Level Run Emphasis (LGRE), (7) High Gray Level Run Emphasis (HGRE), (8) Short Run Low Gray Level Emphasis (SRLGE), (9) Short Run High Gray Level Emphasis (SRHGE), (10) Long Run Low Gray Level Emphasis (LRLGE), and (11) Long Run High Gray Level Emphasis (LRHGE).

In addition to two types of texture features, CAD scheme also computed MRI kinetic features that are divided into two groups namely, computed from the whole breast region and the bilateral difference between left and right breast region. The computed kinetic features include: average enhancement, standard deviation, maximum intensity, the average enhancement of top 10 percent and average enhancement of top 5 percent. Table 7 shows the summary of the computed image features. The similar kinetic features computed from breast MR images have been applied to assist breast cancer diagnosis (i.e., classification between malignant and benign cases) [53] and prediction of breast tumors to neoadjuvant chemotherapy [74].

Table 7: Summary of 38 computed features including texture (GLCM, GLRLM) and kinetic features.

Feature number	Feature group	Description
1-17	GLCM [a]	Contrast, Correlation, Energy, Homogeneity1, Homogeneity2, Entropy, Autocorrelation, Cluster Prominence, Cluster Shade, Cluster Tendency, Dissimilarity, Inverse Variance, Difference Entropy, Maximum Probability, Sum Average, Sum Entropy, Variance
18-28	GLRLM [b]	SRE, LRE, GLN, RLN, RP, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE
29-33	Kinetic for the Background parenchymal area	Average intensity, standard deviation, maximum pixel intensity, average value of top 1%, and average value of top 5% of pixel values
34-38	Kinetic for absolute bilateral difference of BP area	Average intensity, standard deviation, maximum pixel intensity, average value of top 1%, and average value of top 5% of pixel values

[a] GLCM: Gray-Level Co-Occurrence Matrix, [b] GLRLM: Gray-Level Run Length Matrix

3.2.3 Machine Learning Models and Data Analysis

In order to fuse multiple features to yield the improved predictive performance than using single feature, we applied machine learning methods to build the optimal predictive model. Although many different machine learning models have been investigated and applied in CAD of medical images, we selected two classification models or classifiers that represent different machine or statistical learning concepts aiming to predict pCR of

each case underwent neoadjuvant chemotherapy. These two models are 1) fine gaussian support vector machine (FGSVM), and 2) weighted *K*- nearest-neighbor (KNN). In these two learning models, the fine gaussian SVM builds a single global optimization function to make the finely-detained distinctions between two classes using Gaussian kernel scale; while the weighted KNN model builds many local regions based optimization functions to distinguish two targeted classes based on the similarity of the local neighbors.

However, the image dataset used in this study is unbalanced between the two categories of pCR and non-pCR (53 vs. 101 cases). In order to reduce case election bias and more effectively train machine learning models using the balanced image datasets in two categories, we applied a computer program to randomly select 53 non-pCR cases from the dataset to create a new sub-dataset of 106 cases (53 pCR cases vs. 53 non-pCR cases). This computer programs controlled random process is repeated 50 times to generate 50 sub-datasets. Using each sub-dataset of 106 cases, we applied a ten-fold cross validation method to train and test performance of each machine learning model. This training and cross-validation process was repeated 50 times by using 50 sub-datasets to assess robustness of the model performance.

In each ten-fold cross validation, 53 pCR and 53 pCR cases were randomly and individually segmented into 10 subgroups. In each training and testing, 9 subgroups were used to train the classification models. The remaining subgroup of cases was used as an independent dataset to test the classification model and generate the prediction scores. After completing this 10-fold cross validation cycle, each case in this dataset has a classification ranging from 0 to 1. The higher score indicates the higher likelihood or

probability of the case being pCR. Using the classification scores, a publicly available receiver operating characteristics (ROC) curve fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/,University of Chicago) was applied to compute the area under ROC (AUC), which is used as an evaluation index to assess and compare performance of each machine learning model. Finally, mean and 95% confidence interval of each machine learning model using all 50 datasets were computed.

3.3 Result

CAD scheme used the new 50 generated datasets based on statistical bootstrapping concept to build 2 different classification methods including fine gaussian SVM (FGSVM) and Weighted KNN models using all 38 features. Each dataset was used to build different machine learning classifier to predict the result of pathology. Table 8 illustrate the 95% confidence interval (CI) value of the computed AUC value from the ROC curve of these classification. It shows that using FGSVM or WKNN both gives high 95% CI and mean value.

Table 8: Comparison of 95% confidence interval (CI) and mean value of computed AUC using different classification methods.

Classification method	AUC (95% CI, mean value)
FGSVM	[0.66, 0.7], 0.68
WKNN	[0.68, 0.71], 0.69

Table 9 demonstrate the highest accuracy, sensitivity and specificity computed from the two classification methods using all feature pools. Results show that using fine gaussian SVM as the classification to train and test the dataset, yield the highest accuracy,

positive predictive value (PPV) and negative predictive value (NPV)of 62% and 80%, respectively.

Table 9: Comparison of three parameters of accuracy, sensitivity and specificity in three different classification method.

Classification method	Accuracy	PPV	NPV
Fine Gaussian SVM	67%	62%	80%
WKNN (k=10)	59%	62%	58%

3.4 Discussion

Neoadjuvant chemotherapy is the first step for treating breast cancer patients with the aim of shrinking larger tumor [44, 77-78]. A significant fraction of the patients' response completely to the chemotherapy and the tumors are vanished, while some patients only have partial or no response to the neoadjuvant chemotherapy. Although it is understandable for patients who have complete response shown in breast MRI acquired post-therapy to not undergo for surgery, but the complete disappearance of the tumor or malignant cells can only be determined by a comprehensive pathology examination. The pathological analysis will be performed on many breast tissue specimens that was removed by the surgery to examine whether there are the remaining malignant tumor cells. Since patients who have the pCR reported higher survival rate [79], prediction of pCR without aggressive surgery is a research topic attracted broad interest in current clinical research field of breast cancer. Thus, identifying non-surgical or non-invasive imaging markers to predict pCR play important role to omit the unnecessary and aggressive surgery for the breast cancer patients with pCR to the neoadjuvant chemotherapy.

In this study, we developed and tested a new CAD scheme to investigate the association of MRI image features computed from the DCE-MRI scans performed after neoadjuvant chemotherapy in predicting the pCR of breast cancer. We demonstrated that using texture features along with the kinetic image feature computed from breast MRI after performing neoadjuvant chemotherapy carry useful information for predicting the pathologic response. It was shown in this study that fine gaussian support vector machine classifier can predict pCR with higher accuracy compared to other classifiers. To the best of our knowledge, the prediction of pCR for patients receiving neoadjuvant chemotherapy from the breast MR images using the fusion of kinetic and texture features has not yet been investigated.

However, there are some limitations to our work. First, the number of the dataset is limited, and it is tried to perform our CAD scheme on a larger dataset to validate the performance of our CAD scheme. Second, different definition of pathological complete response should be added to investigate the different subtype of the tumors. In conclusion, we demonstrated that image features including kinetic and texture features extracted from breast MR images acquired after performing neoadjuvant chemotherapy can help in predicting pathological response to neoadjuvant chemotherapy. Finding of this study can be helpful in selecting patients undergoing surgery for the pathologic analysis.

Chapter 4: Implementation of a Computer-aided Detection Tool for Quantification of Intracranial Radiologic Markers on Brain CT Images

4.1 Introduction

Spontaneous, non-traumatic aneurysmal subarachnoid hemorrhage (aSAH) is most commonly due to ruptured cerebral aneurysm with an annual incidence of ~30,000 per annum in the US and affects middle-aged population with higher incidence among women and is associated with significant morbidity in survivors [80, 81]. This category of hemorrhagic stroke results in high individual and societal socioeconomic burden [82, 83]. It is important for the clinicians to develop and implement an optimal personalized treatment and rehabilitation strategy for respective patients to help them recover quickly and also counsel families accurately regarding futility of care in patients with severe/near fatal aSAH. For this purpose, imaging plays an important role [84, 85] in current clinical practice. This calls for development of more effective radiological imaging based clinical markers that have higher discriminatory power to predict or assess prognosis of the stroke (aSAH) patients. However, due to the large number of images (CT or MRI), as well as the use of the subjective assessment methods, there are few quantitative studies in aSAH population [86, 87], which is not only labor-intensive (time-consuming), but also inconsistent due to the large inter-reader variability.

In order to more accurately and quantitatively predict or assess the prognosis of aneurysmal subarachnoid hemorrhage patients, we developed and tested a new interactive CAD tool to automatically detect, segment and quantify two radiological image markers

namely, the volume and distribution of brain hemorrhage and ventricular cerebrospinal fluid regions depicting on brain CT images. The same interactive CAD schemes have been developed and tested in our previous studies for different types of medical images including mammograms [88 - 90], breast MRI [74], abdominal CT images [91]. In this study, our CAD scheme segments brain skull and assigns pixels into normal brain tissue, blood and fluid. To further increase accuracy in quantification of the image markers, we designed and implemented a graphic user interface (GUI) into the CAD scheme, which allows users to visually examine segmentation results and guide CAD scheme to automatically correct errors. Thus, unlike the qualitative features that were assessed and provided by the clinicians (e.g., radiologists) with established neuroradiological scales, namely Mod Rotterdam scale and Helsinki scale [92, 93], using our CAD enables to extract and compute quantitative image features. The overall goal of this study is to provide the clinicians a new computer-assisted tool to accurately and robustly measure these two clinical markers in the clinical settings using the brain CT images under potentially large variation of CT image characteristics and/or noise levels.

Currently, this new interactive CAD has now been installed in the clinical and research settings in Department of Neurology, the University of Oklahoma Health Science Center (OUHSC) to conduct further clinical tests and studies. To report the study of developing this new interactive CAD scheme and results of testing the feasibility of using this interactive CAD scheme in our medical imaging laboratory and OUHSC clinical site, the organization of this Chapter is as follows: Section II and III describe the image dataset and the proposed CAD scheme applied over image datasets used in this study, respectively. Section IV summarizes our experimental results and Section V

discusses the observation and conclusions that we can make from our study results, respectively.

4.2 Materials

An image dataset of brain CT images including 96 patients who met the inclusion criteria was assembled. Each patient had the series images of axial view with image slides number between 30-60 slides. Of the 96 patients, 65 (68%) were female and 31 (32%) were male. In this dataset, mean and standard deviation of the patients' age are 51.8±12.4 and 51.2±13.2 for female and male, respectively, indicating that the majority of patients with brain hemorrhage were selected in this study are relatively older, i.e., >50 years old. Among our patients, different discharge disposition such as home, rehab and expired had been recorded.

4.3 Methods

We developed and tested a new CAD scheme to automatically segment brain tissue by removing brain skull and detecting brain hemorrhage and ventricular cerebrospinal fluid regions depicting on brain CT images. Once a case of brain CT examination is loaded into the program, an automated process steps are applied. There are five main image processing steps including: (1) applying region growing algorithm with adaptively adjusted threshold to remove brain skull and segment brain regions, (2) using thresholding algorithm to detect normal brain tissue, blood and fluid regions, (3) examine the segmentation result and applying the needed correction; and (4) computation of image features. The schematic overview of the entire segmentation workflow is presented in

Figure 11 and the details of each image processing steps are described in the following sections. Additionally, to avoid a “black-box” type approach and increase the confidence of the users to accept or consider CAD results, we also developed and implemented an easy-to-use graphic user interface (GUI), which make this CAD scheme an interactive CAD scheme.

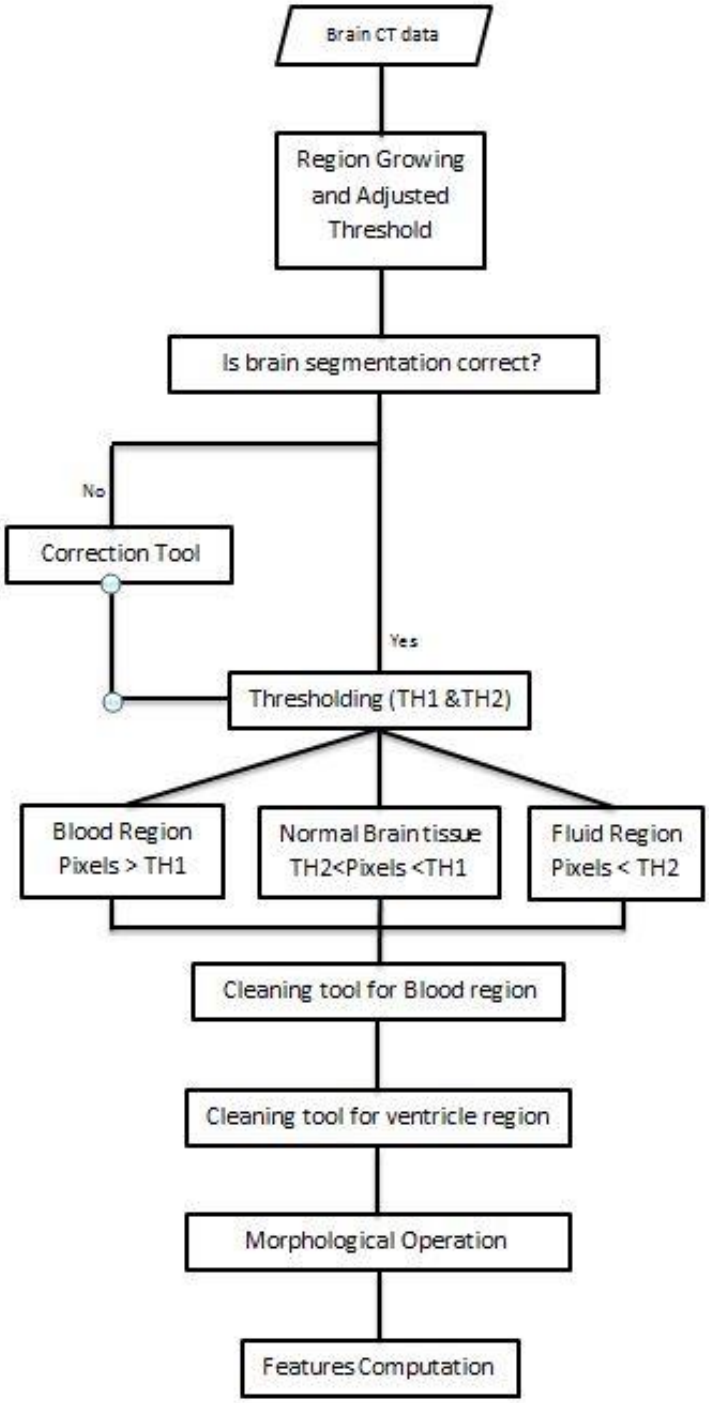


Figure 11: Flow chart of subsequent steps of the brain segmentation

4.3.1. Brain region segmentation

Brain region segmentation is an important step for the CAD scheme to automatically remove brain skull and segment brain tissue. Our CAD scheme first asks the user to create a seed point on brain tissue and apply the region growing algorithm to segment brain region and remove brain skull. Our CAD scheme uses a thresholding algorithm to differentiate regions from brain tissue and brain skull.

4.3.2. Normal brain tissue, blood and fluid regions detection

Second, in order to sign each pixel inside the segmented brain region into three categories or classes namely, normal brain tissue, blood and fluid, our CAD scheme applies a thresholding algorithm with two different values (TH1 > TH2) to detect desired regions. Since blood region is more enhanced in brain tissue, pixels with CT number above TH1 are segmented as blood region, while pixels with CT number less than TH2 are segmented as fluid region. Pixels with CT number between TH1 and TH2 are classified as normal brain tissue. In the GUI window, the segmented normal brain tissue, blood and fluid regions are demonstrated in gray, light gray and black color, respectively.

4.3.3 Segmentation examination and correction

Although our CAD scheme enabled to achieve high accuracy in brain and ROI segmentation, errors do happen in a small fractional of image slices (e.g. $\leq 5 - 10\%$). In order to solve this practically operational issue and optimally compensate potential impact of CT image noise variation on different cases, users can visually examine segmentation results slice-by-slice to make sure that the segmentation results and the

correspondingly computed volumes of brain tissue, blood and fluid are satisfactory or acceptable. Figure 12 demonstrates an example of the process of our scheme segmentation. For this purpose, we designed and implemented a number of function buttons for the interactive correction which are listed below.

1) Correction button – In the first step of brain region segmentation, if the segmentation has error (i.e., part of the brain is missing), the user can click a correction button, the scheme automatically corrects the error by mapping the new boundary condition based on the brain segmentation results in the adjacent CT slices.

2) Removing false regions – CAD scheme has a function to allow the users convert the blood or fluid located in the irrelevant regions in to normal brain tissue region. An automatically converting process will be applied to the ROI created by the user.

3) Morphological operations – This function button allows to perform morphological operations only inside the ROIs defined by the user to fill in the holes inside the blood or fluid regions.

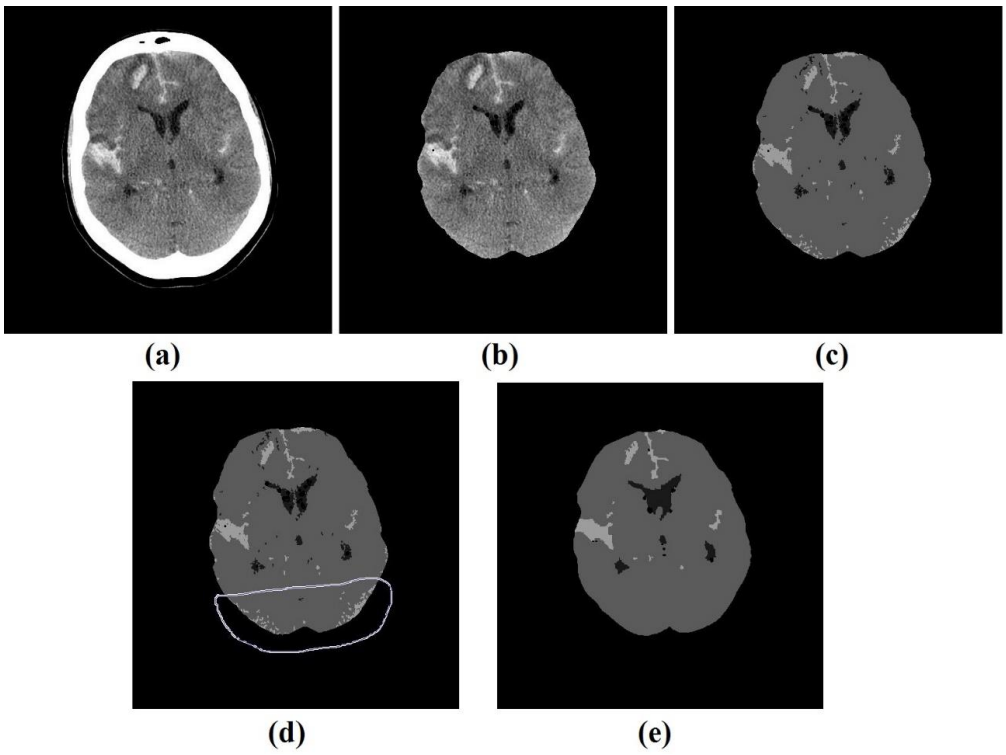


Figure 12: An example of process of our CAD scheme. (a) Original image, (b) Skull removal and brain segmentation (c) Brain normal tissue, Blood and Fluid region detection which can be seen in gray, light gray and dark gray colors, respectively. (d) Removing false regions by creating a boundary around the ROI, (e) Final image ready for feature computation.

4.3.4 Feature computation

Finally, once the user satisfies with the results of each region segmentation, the CAD scheme computes total of 6 image features from the segmented regions, which are volume of total cranial, brain tissue, blood, fluid and ratio of blood volume over total cranial volume, and ratio of fluid volume over total cranial volume. These quantitative image features are displayed in the GUI window and then automatically saved into an Excel datasheet.

4.4 Results

Under the IRB-approved data collection and study protocol, we have retrospectively acquired and assembled an initial testing image dataset involving pre-contrast image series of 96 brain CT scanning cases in this study. Each case was of patients with non-traumatic aSAH of different clinical and radiological grades regarding to the disease severity. We applied our interactive CAD scheme to process dataset and to compute 6 image features as described above. Table 10 summaries minimum, maximum and average of the computed image features. The ratios range from 0.12% to 13.46% in this dataset, which indicates a big variation in the quantitatively assessed blood volumes among different patients in this dataset.

Table 10: Summary of minimum, maximum and average of the quantitative image features.

Features	Minimum	Maximum	Average ± STD
Brain Volume	958.62	1492.53	1156.2
Blood Volume	0.92	94.92	18.11±18.29
Ventricles Volume	5.52	91.63	27.67±16.28
Tissue Volume	1079.58	1629.44	1296.29±139.03
Ratio of Blood Volume over Brain Volume	0.12%	13.46%	2.4%±2.57%
Ratio of Blood Volume over Ventricles Volume	0.82%	9.47%	3.51%±1.81%

Figure 13 shows an example of the interactive region segmentation steps and results. An original CT slice is displayed on the left. The image in the middle shows the result generated by the CAD scheme in the first automatic processing step, which shows the errors including missing a part of brain region (on the top) and detection of non or irrelevant blood ROIs or pixels due to the unavoidable CT image noise when using the regular threshold on CT number. After taking the interactive correction steps, the right image shows the final segmentation results in segmentation and quantification of brain region, as well as blood and fluid regions. The bottom part of the figure also shows the updated computation results.

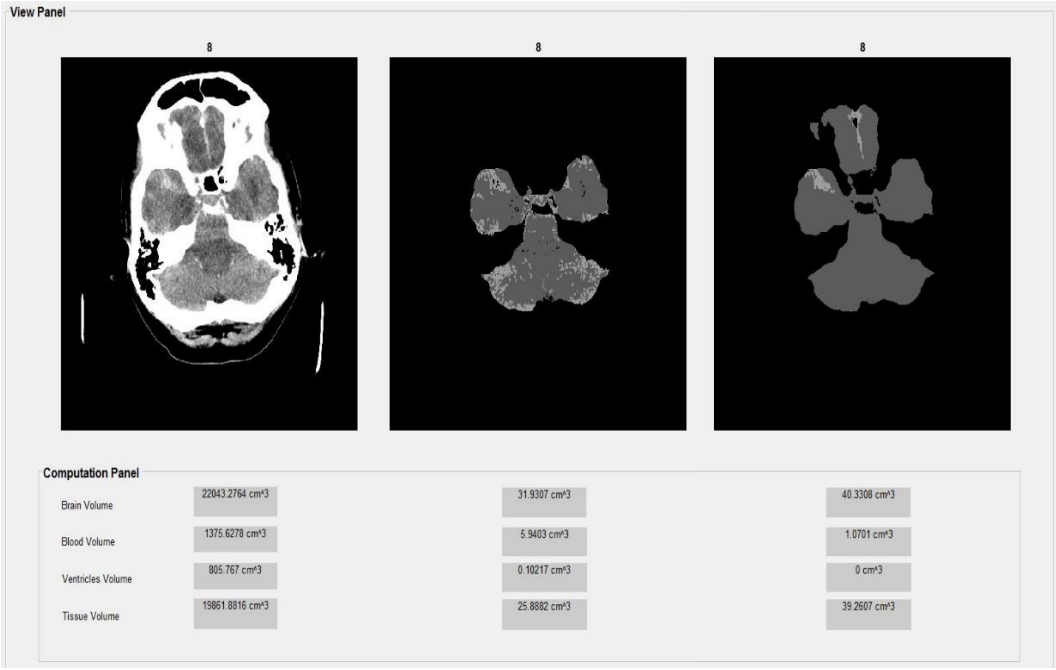


Figure 13: An example of brain segmentation using the correction tool to correct missing brain tissue area. Left image shows the original CT image, middle image is the segmented brain region missing one top bran tissue region and right image shows corrected image with both automatically and manually for all segmented regions. At the bottom of each image, 4 features were computed. Left column is related to total whole

series, middle column is related to middle image and right column is the features for corrected image. It can be seen by every changes value of computed features differ.

We then analyzed the association between these features and the neuroradiological scales, namely Mod Rotterdam scale and Helsinki scale, commonly used in current clinical practice to assess severity and prognosis of stroke patients with aSAH. Table 11 and Figure 14 demonstrate the relationship or correlation coefficient between the computed image features (or segmented volumes) and the validated clinical evaluation indices (Mod Rotterdam and Helsinki sales). Among these relationships, ratio of blood volume over brain volume shows the highest correlation value with both Mod Rotterdam scale and Helsinki scale of 0.64 and 0.62, respectively.

Specifically, the fit of the linear regression line equation for ratio of blood volume over brain volume and Mod Rotterdam scale was

$$y = 0.0146x + 0.0032$$

and the linear regression line equation for ratio of blood volume over brain volume and Helsinki scale was

$$y = 0.0067x + 0.0006$$

with $R^2 = 0.41$ and $R^2 = 0.39$, respectively. Therefore, for a computed ratio of the blood volume over brain volume of 0.04, Mod Rotterdam scale and Helsinki scale were 2 and 4, respectively.

Table 11: Correlation Coefficient between the computed image features and the clinical features.

	Helsinki Scale

Clinical Features **Computed Image Features**	Mod Rotterdam Scale	
Brain volume	-0.04	-0.02
Blood Volume	0.61	0.6
Ventricles Volume	-0.04	-0.02
Tissue Volume	0.01	0.03
Ratio of Blood Volume over Brain Volume	0.64	0.62
Ratio of Ventricles Volume over Brain Volume	-0.03	-0.02

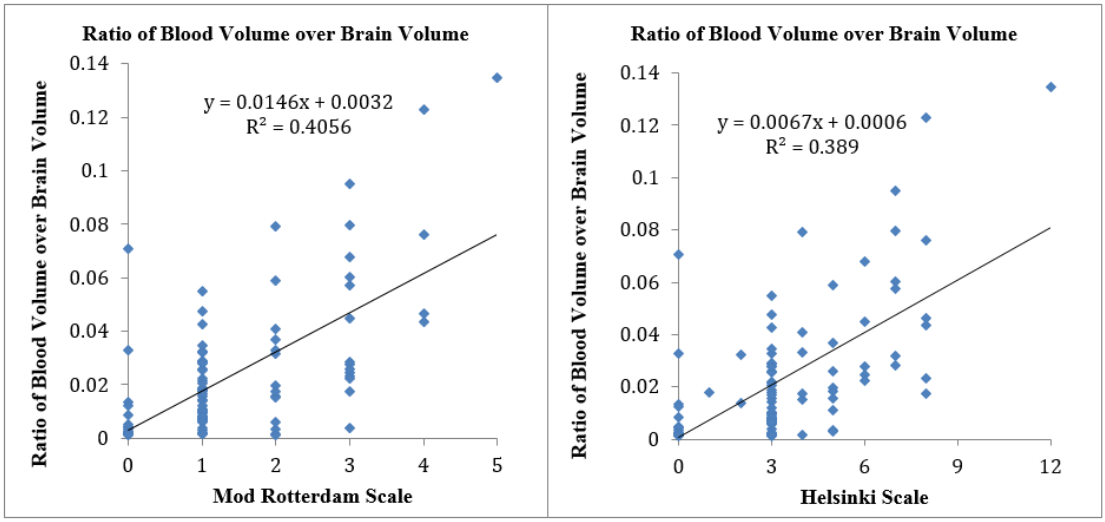


Figure 14. Scatter plots of ratio of blood volume over brain volume with Mod Rotterdam scale and Helsinki scale. Solid line is the linear regression line.

4.5 Discussion

Based on the imaging informatics concept, we in this study developed and demonstrated a new interactive CAD scheme/tool to help detect, segment and quantify brain volume, hemorrhage and CSF regions depicted on cranial area of brain CT images.

Our scheme cannot only automatically segment brain volume, the blood and CSF regions with a relatively high accuracy and computational efficiency, it also provides the unique interactive functions for the clinical researchers or users to conveniently examine and make correction (if necessary). This new ICAD scheme has been repeatedly tested by research assistants in our medical imaging (or CAD) laboratory. The testing results have been visually examined. Although CT image quality (signal-to-noise ratios) vary greatly among the brain CT image dataset, we found that using this interactive CAD scheme and GUI tool enabled users to adaptively process brain CT images and yield satisfactory results for the purpose of accurately and robustly segmenting and quantifying brain hemorrhage and ventricular cerebrospinal fluid depicting on the non-contrasted brain CT images. The preliminary feedback from the clinical researchers has also be encouraging, which indicated that using this new ICAD tool was much more user friendly and less cumbersome than other "black-box" type automated tools that have been previously tested for this type of studies to quantify various intracranial components or radiologic markers.

Thus, our study convinced our hypothesis that for this type of task aiming to accurately quantify two image based clinical or radiological markers using brain CT images of the stroke patients with spontaneous aSAH, accuracy and users' confidence on the CAD-generated quantitative data is the most important if the such type of CAD scheme can be eventually accepted in the clinical practice to quantify image based markers to predict or assess prognosis of the aSAH patients. The interface of the interactive functions of our ICAD scheme have been modified or optimized based on the feedback of the clinical and research users. The scheme has now been installed in the

clinical and research settings and used by medical students or clinical researchers for conducting the related clinical studies to investigate several homeostatic derangements that occur after aSAH and correlate them with the clinical outcome.

After installing this interactive CAD scheme and GUI tool in the Department of Neurology of OHHSC, the clinical researchers have used this ICAD scheme to evaluate the relationship between glycemic gap (GG) determined stress-induced hyperglycemia (SIH), modified Graeb score (mGS) and estimated intracranial blood and cerebrospinal fluid volumes, which have reported encouraging results in two papers recently published in two clinical oriented journals [94,95]. Specifically, volume of blood, CSF, and brain tissue (in ml) is estimated using this software program. Intracranial volume is determined as a combined volume of the above three parameters. Intracranial blood volume is expressed as a ratio of (blood volume to intracranial volume) × 1000. Such a ratio is used to obtain amount of intracranial bleeding in reference to respective individual's intracranial volume. Similarly, total of blood volume and CSF estimates fluid volume and is expressed as (blood volume + CSF to intracranial volume) × 1000 to provide quantitative estimation of hydrocephalus.

In the clinical study of OUHSC, the researchers retrospectively assembled a brain CT image dataset involving 187 patients over a 5-year period, which meet the inclusion criteria for the study. Among them, 38 patients (20.3%) expired during hospitalization, while 9 patients (4.8%) were discharged to an inpatient hospice facility. A GG of 26.7 mg/dl was considered as the optimum threshold for SIH which resulted of ninety-four (50.3%) among the study cohort that had SIH. Baseline characteristics including patient demographics are enumerated in Table 12.

Table 12: Patients demographics information.

Radiological feature	SIH (GG ≥ 26.7 mg/dl)	Mean/median (in ml)	95% CI/IQR	p value
Intracranial blood volume (in ml per 1000 ml cranial cavity)	Yes	39.6	33.6–45.5	0.0002
	No	25.3	20.6-29.9	
Intracranial blood and CSF volume (in ml per 1000 ml cranial cavity)	Yes	79.1	71.9-86.2	0.027
	No	63.1	56.2-70.1	
Modified Graeb score	Yes	4.0	2.0-7.0	0.002
	No	2.0	0.0–6.0	

Data analysis results indicated using this GG threshold, patients with SIH had 14.3 ml/1000 ml more intracranial blood volume and higher mGS as compared to those without SIH [39.6 ml (95% confidence interval, CI, 33.6 to 45.5) vs. 25.3 ml (95% CI 20.6 to 29.9), p = 0.0002] (Figure 15a, Table 12). Also, patients with third ventricular blood on admission CT scan were more likely to develop SIH [67/118 (56.8%) vs. 27/69 (39.1%), p = 0.023].

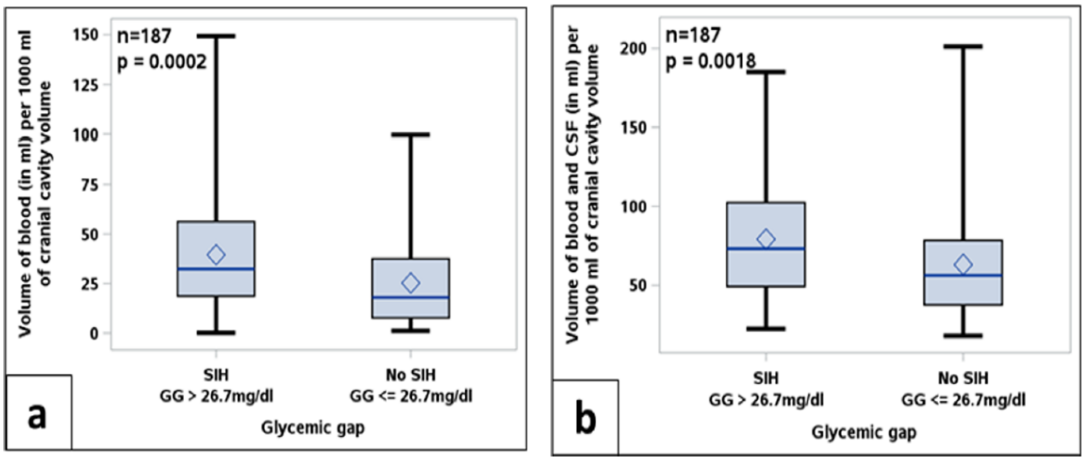


Figure 15: Box plots show significant difference among patient developing stress-induced hyperglycemia regarding to their respective intracranial hemorrhage volume (a) and degree of hydrocephalus (b) using quantitative estimation.

Similarly, those with SIH had, on an average, 16/1000 ml of more intracranial blood + CSF volumes as compared to those without SIH [79.1 ml (95% CI 71.9 to 86.2) vs. 63.1 ml (95% CI 56.2 to 70.1), p = 0.027] (Fig. 15b; Table 12). Also, linear regression analyses showed strong relationship between volume of intracranial blood and intracranial blood + CSF with mGS. Hence, the present study, using unbiased SIH definition and objective CT scan parameters, reports "dose-dependent" radiological features resulting in SIH. Such findings allude to a brain injury-stress response-neuroendocrine axis in etiopathogenesis of SIH.

Based on our initial tests, our next short-term goal is to expand the retrospective database to include over 300 stroke patients with aSAH. We are now planning to apply this new ICAD scheme to process all of these cases and compute the cerebral hemorrhage and/or ventricular cerebrospinal fluid related image (clinical or radiologic) markers. The

detailed statistical data analysis study to assess the association between these image-based markers and prognosis of the aSAH patients will follow. Although reporting the comprehensive data analysis of this database of over 300 aSAH patients is beyond the scope of this technology development study, this technology development based study is an important step and successfully demonstrated its feasibility or clinical utility in order to make the contribution of helping clinicians more effectively conduct quantitative studies to investigate and address the important clinical issues or questions in the future.

Chapter 5: Computer-Aided Antibody Screening for IHC Assay Development

5.1 Introduction

Immunohistochemistry (IHC) is an important and commonly used method for cancer diagnosis in pathology, which is used to visualize certain antigen (protein) as the target in cellular (tissue) samples using monoclonal or polyclonal antibodies. In most cases, the antibody is conjugated to an enzyme, which catalyzes a color-produced reaction that enables the detection of the antigen. Alternatively, the antibody can be conjugated to a fluorochrome to detect the presence or absence of the marker.

Thus, identifying the right antibody that can conjugate specifically to the targeted disease is the most important task in the whole IHC assay development process. To achieve this, a large amount of antibodies need to be screened, which is conventionally performed by pathologists through manual slides reviewing. Usually, slides stained by 3,3´-Diaminobenzidine (DAB) with hemotoxylin (HTX) counter stain are used, where the presence of specific DAB stain indicates the desired reaction of the antibody. It is known that 30%~50% of the antibodies under screening can be non-reactive to the target, resulting in slides with little DAB stain (i.e, negative slide).

The purpose of this study is to develop a new CAD scheme to automatically identify the non-negative slides and thus reduce the pathologists' manual reading time, which aims to eventually improve the efficiency of antibody development. We focused on using

both low and high resolution digital pathology images to extract image features in a two-step scheme to optimize algorithm efficiency and accuracy, especially for the challenging cases where the desired specific DAB stain was very weak or sparse on the slide. We also experimented with different machine learning algorithms, including hand-crafted and convolutional neural network (CNN) based feature extractions. In addition, the algorithm was integrated to a web-based image management platform to facilitate easy deployment to the clinical users.

5.2 Materials

Two digital pathology image datasets with 285 and 279 slides were retrospectively assembled as training and testing dataset, respectively. Each dataset was divided into two groups namely, negative vs. non-negative groups. The negative group included images with little DAB stain and non-negative group included images with partial or pure background DAB stain. The training dataset included 126 negative and 159 non-negative slides, while the testing dataset included 130 negative and 149 non-negative slides, respectively. Figure 16 demonstrates several sections of the images namely, the field-of-view (FOV) examples of negative and non-negative slides, which shows that the problem is not trivial, as DAB stain can present in negative slide (Fig.16 (b)), and specific DAB stain can be very sparse in non-negative slide (Fig.16 (e)).

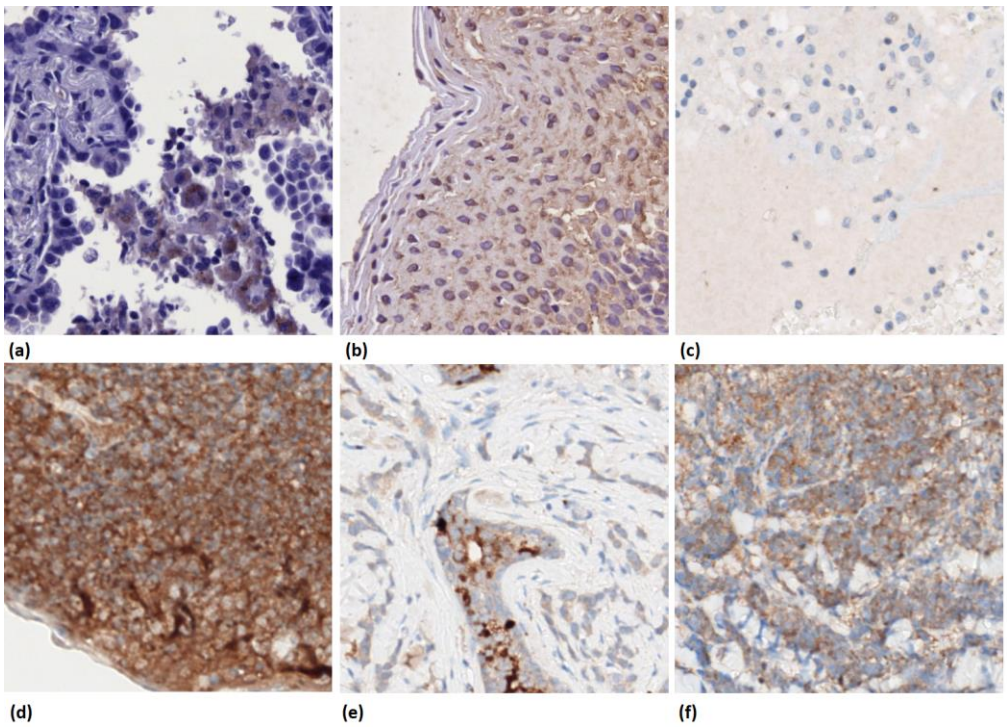


Figure 16: Examples of FOV images (a-c) from negative and (d-f) non-negative slide.

5.3 Methods

In the study, our CAD scheme first applied color deconvolution to the low resolution whole slide image and extracted histogram based image features. The color deconvolution method developed by Nie et al [96] was used to unmix the digital pathology image into single stain images. As the result, three new images (i.e., HTX, DAB and total Absorbance of all the stains) were generated, which were used for feature computation. Different machine learning algorithms including linear supported vector machine, (LSVM), k-nearest neighbors (KNN) were built using the training dataset through a ten-fold validation method. Next, as the initial classification process, the trained models were tested over the testing dataset. The classification results showed that LSVM

achieved the highest area under ROC curve (AUROC). Therefore, we used the LSVM score to derive the cutoff points for identifying slides that need further analysis. The lower- and upper-cutoff points were defined such that there were no misclassified slides from the training dataset with a LSVM score that is lower(higher) than the lower(upper)-cutoff points.

In the next step, whole slide images that needed further analysis were divided into non-overlapping tiles at high resolution. Color deconvolution was applied to each tile and more sophisticated image features were extracted. We experimented with histogram based features, as well as features derived from the intermediate layers of pre-trained convolutional neural network (CNN), such as AlexNet. Based on these extracted features, whole slide image representation was derived utilizing a Bag-of-Words (BoW) model. Specifically, image tiles from the training dataset were clustered into a predefined number of clusters; and the centers of the clusters collectively form a visual words vocabulary. For each tile in a whole slide image, a visual word was assigned by identifying the cluster that the tile belongs to. Thus, each whole slide image can be represented by the histogram of the occurrence of each visual word in the vocabulary. Finally, the second classifier is trained using the BoW whole slide image representation to perform the negative vs. non-negative prediction. The schematic overview of the entire classification workflow is presented in Figure 17.

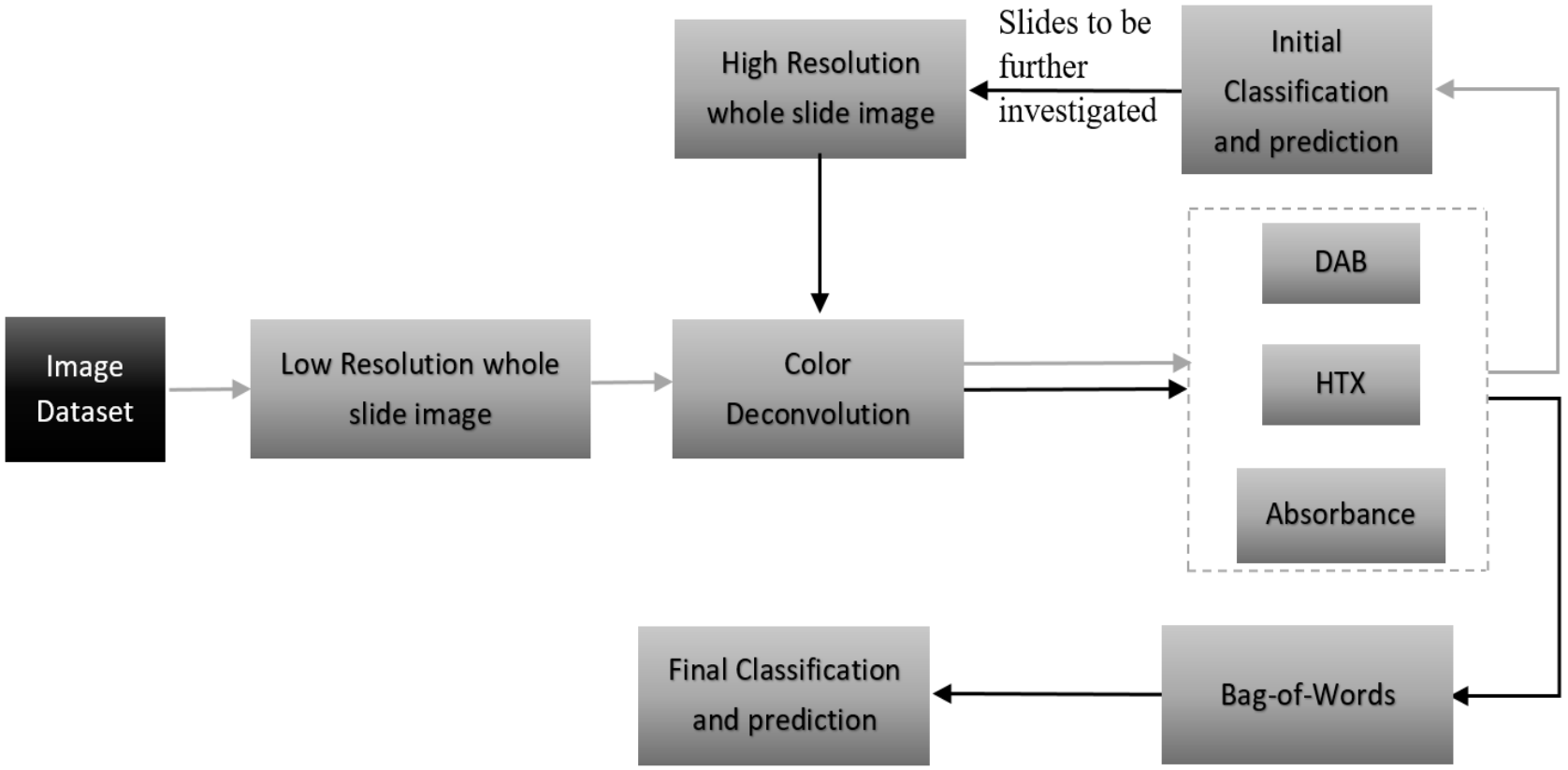


Figure 17. Flow chart of two-stage classification scheme.

5.4 Results

Figure 18 shows several example results of applying the color deconvolution method on the low resolution whole slide image.

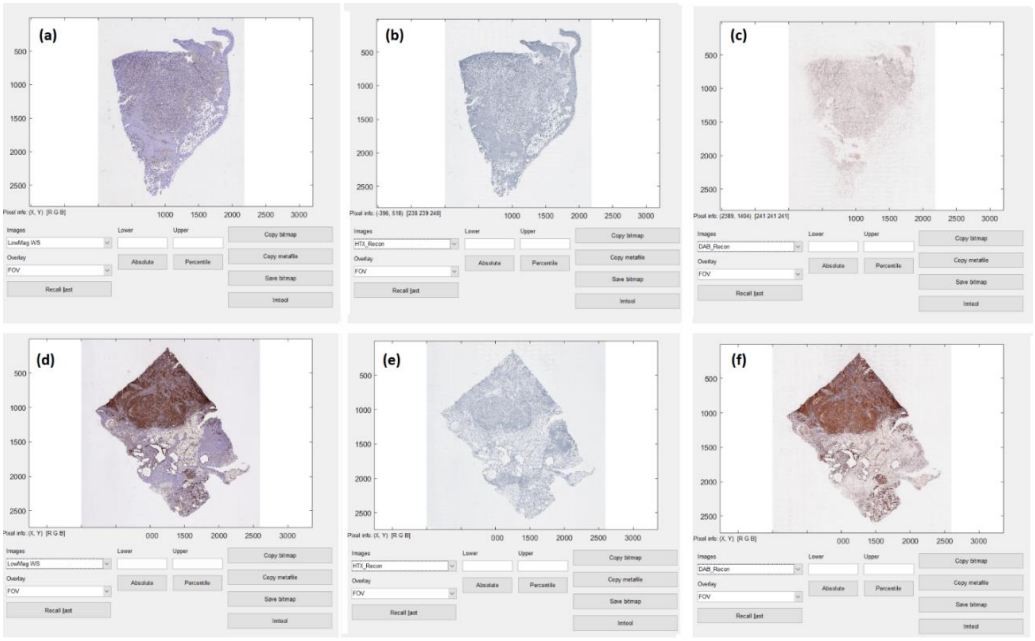


Figure 18: The graphical user interface developed for showing color deconvolution results. a) low resolution negative image, b) HTX generated from the negative image, c) DAB generated from the negative image, d) low resolution non-negative image, e) HTX generated from the non-negative image and f) DAB generated from the non-negative image.

For the initial classification, different machine learning algorithms were applied to build prediction or classification models using the training dataset. The trained models were then tested by the independent testing dataset. Table 13 shows a list of machine learning classifiers and their performance on the testing dataset. Among them, linear

SVM shows higher accuracy of 82% with AUROC of 0.86. The sensitivity and specificity of the linear SVM were 80% and 84%, respectively.

Table 13: Comparison of prediction performance using different machine learning models.

Machine Learning Model	Accuracy	Sensitivity	Specificity
Linear SVM	82%	80%	84%
Fine Tree	70%	68%	72%
Fine KNN (K=1)	75%	74%	76%
Medium KNN (K=10)	75%	68%	85%
RusBoosted Tree	68%	65%	70%
Boosted Tree	57%	55%	57%

Figure 19 shows the histogram plot of the initial classification result using linear SVM model for the negative and non-negative slides of the training dataset. Results indicated that the cutoff points are -1.5 and 2.75 for negative and non-negative slide, respectively. Any slide with a linear SVM score between the two cutoff points was identified as a slide which needed further investigation.

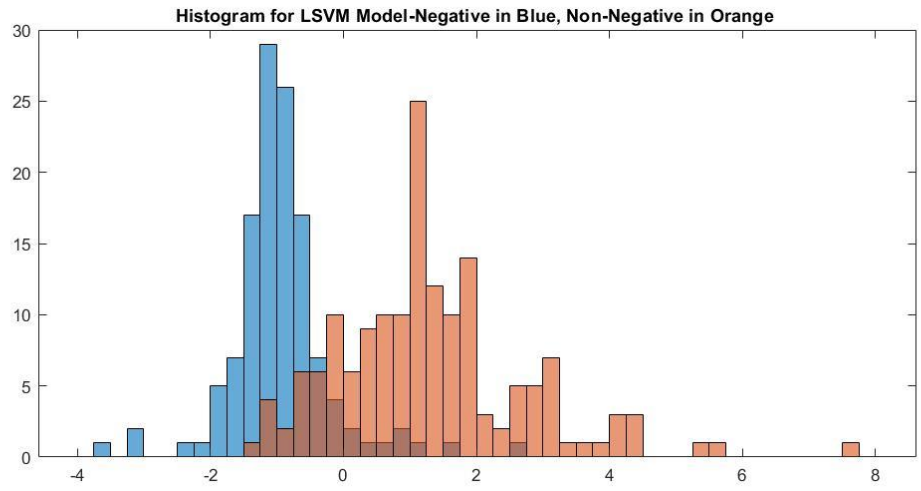


Figure 19: The histogram plot of the linear SVM scores in the initial classification. Negative and non-negative slides are shown as blue and orange, respectively.

From the training dataset out of 285 cases, and from testing dataset out of 279 cases, 244 and 218 cases were between the two cutoff points, respectively. In our scheme, a total of 180 tiles of 500x500 pixels were generated from each high resolution whole image to apply the final classification. The tile selection was performed by first computing the mean DAB intensity value of each tile. Then, different percentiles of the mean DAB intensity values were calculated such that half of which were evenly distributed between the 80th~99th percentiles, and others were evenly distributed between the 20th~80th percentiles. For each selected percentile, a same number of tiles with mean DAB intensity value closest to the percentile were selected, resulting in the final 180 tiles to represent the whole slide image, and the majority of which had relatively high mean DAB intensities. The rationale behind this tile selection approach is based on the assumption that tiles with relatively higher mean DAB intensity values in each slide are more relevant

for differentiating negative vs. non-negative slides, which was found true though our experiments.

The stain intensity histogram based features and features derived from the intermediate layer of the pre-trained AlexNet network were extracted for each tile to derive the BoW representation for the whole slide image. Different number of clusters including 30, 35, 40, 45 and 50 were tested to evaluate the prediction performance. Using the training dataset, 21 different machine learning classifiers were built, trained and tested over testing dataset, including Fine K-Nearest Neighbor (KNN), Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, Weight KNN, Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, Coarse Gaussian SVM, Ensemble Boosted Tree, Ensemble Rusboosted Tree, Fine Tree, Medium Tree, Coarse Tree, Ensemble Subspace Discriminant, Ensemble Subspace, Ensemble Bagged Tree and Linear Discriminant. Table 14 and 15 show a performance comparison of different number of clusters using the trained models using the histogram-based features and neural network generated features, respectively. For each cluster number, the results for the best classifier with accuracy value above 70% are reported.

Table 14: Comparison of the best classification performance using different number of clusters over histogram based features.

Machine Learning Model - # of clusters	Accuracy	Sensitivity	Specificity
EBGT* - 30	**74%**	**83%**	**66%**
CGSVM - 35**	**76%**	**88%**	**66%**
EBGT - 40	**80%**	**85%**	**74%**
CGSVM - 45	**77%**	**91%**	**67%**
Ensemble Subspace KNN - 50	**80%**	**82%**	**76%**

Table 15: Comparison of the best classification performance using different number of clusters over neural network based features.

Machine Learning Model - # of clusters	Accuracy	Sensitivity	Specificity
Ensemble Subspace Dis - 30	81%	84%	77%
Ensemble Rusboosted Tree - 35	84%	80%	89%
EBGT- 40	83%	91%	73%
Ensemble Subspace Dis - 50	81%	82%	79%

Results from Table 14 indicate that although using Ensemble Bagged Tree (EBGT) classifier with number of clusters of 40 achieved higher accuracy, using Coarse Gaussian SVM (CGSVM) classifier yielded the highest sensitivity of 91% and AUROC curve value of 0.88. While EBGT showed better accuracy for using the histogram-based features, Ensemble Rusboosted Tree classifier showed better performance for using neural network generated features. Also using EBGT classifier for neural network generated features achieved overall highest performance in accuracy, sensitivity and

AUROC with the value of 83% and 91% and 0.91, respectively. It is worth mentioning that, superior sensitivity is preferred for this particular application, because non-negative slides identified by the CAD scheme will be sent to pathologist for further review. Thus, including a small number of misclassified negative slides is more acceptable than missing non-negative slides that may contain important information regarding reactive antibodies.

5.5 Discussion

Primary antibody development for IHC assay is to identify the antibody that is conjugated specifically to the target antigen. For this purpose, the first question to answer is whether the antibody under study is conjugated to the tissue in general regardless of the staining specificity, which is the problem addressed in this study. In comparison to the widely known IHC scoring applications for clinical diagnosis, such as Ki67 and human epidermal growth factor receptor 2 (Her2), etc., where a quantified score needs to be derived through calculating the percentage of positively stained tumor cells, our study and CAD scheme focus on differentiating reactive vs. non-reactive antibodies at whole slide level. Therefore, our focus is naturally formulated as a binary classification problem namely, differentiating non-negative vs. negative slide. Under such formulation, the ground truth available to us is the whole slide level labels, which lead to our strategy of employing slide and tile level image features to derive the whole slide representation for building the classification model. Cell level image features, although potentially more descriptive in local regions, are not considered in this explorative work due to the fact that cell level ground truth is not well defined for this general application. The experiment

results demonstrated that the proposed CAD scheme achieved encouraging results without the need of computationally expensive cell level analysis.

On the other hand, tile selection seems to play an important role in the proposed scheme. In fact, we examined the scheme using all the tiles from each whole slide image, which showed inferior performance in classifying the slides comparing to using only 180 tiles. This could be due to the fact that using all the tiles from a whole slide image results in employing too many tiles with sparse DAB stains in non-negative slide, which often have similar appearances as the tiles in negative slides. Using presented tile selection criteria in this study, better classification performance was achieved, which confirmed our assumption that tiles with relatively higher DAB intensity level were more representative for identifying reactive vs. non-reactive antibodies through whole slide assessment.

However, despite the encouraging results, this is a preliminary study with several limitations. First, this study used a small image dataset which may not sufficiently represent the image diversity in this application. Second, by finding that using all the tiles in the whole may not be the optimal approach. Thus, there exits great room for future investigation on tiles selection to derive a better representation of the whole slide. Moreover, CNN based features were explored as is, more research work is needed to design a training scheme that enables fine tuning the network parameters to further improve the performance for this particular application. Our next step includes investigating the non-negative slide and automatically excluding pure background staining cases, which will further reduce the manual reviewing time by pathologists.

Chapter 6: Developing a CAD-based Slide Label Reader to Improve Efficacy of Storing Digital Pathology Images

6.1. Background

The pathology discipline has its basis in over 150 years old tradition of using microscopes to review and diagnose tissue samples. Tissue samples are collected by surgeons or other clinicians and sent to the pathology laboratory for analysis. The largest hospital in each county is equipped with a pathology laboratory that receives tissue specimens or bodily fluid from other departments or hospitals. The technicians in pathology laboratory processes the material and creates glass slides with pieces of the specimens and finally, a pathologist reviews the content on the glass slide and dictates a report that is sent back to the referring physician.

The typical tissue specimens arrive at the pathology laboratory in a plastic box filled with formalin. A pathologist or laboratory technician performs a grossing examination of the specimen. This examination consists of measuring, sketching, and sometimes taking photos of the specimen. Small pieces of tissue are cut out for further analysis, see Figure 20(up left). These pieces are further processed in a sequence of chemical solutions by an automatic processing machine and then embedded into paraffin blocks (Fig. 20 up right). From these paraffin blocks, small sections of tissue are sliced using a microtome. The microtome enables to create micrometer thin sections, which are placed on warm glass slides where it will be immediately fixed (Fig. 20 down left). At this stage, the resulting sections are almost transparent. In order to reveal the tissue structures, the glass slides are stained with different colors that attach to different types

of structures. For each specimen, this process can result in a large set of glass slides (Fig. 20 down right).

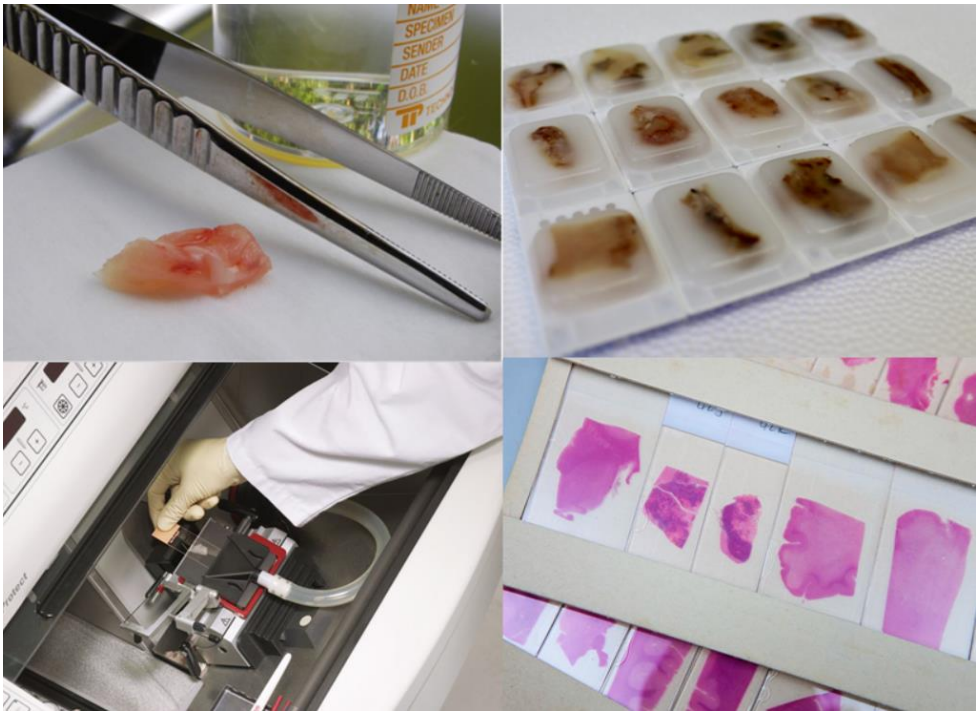


Figure 20. The process of producing glass slides. (up left) Small pieces of tissue are selected for processing. (up right) Pieces of tissue embedded in paraffin block. (down left) Paraffin block are cut with a cryostat microtome and being prepared for staining. (down right) Tray of glass slides ready for review.

Finally, the pathologist reviews the content of the glass slides with a microscope and dictates a report. This work is typically organized as a sequential production line in multiple steps, consisting of different kinds of processing and selection steps as in Figure 21. If the patient has given written consent, the glass slides and the embedded blocks are stored for future research and follow-up purposes.

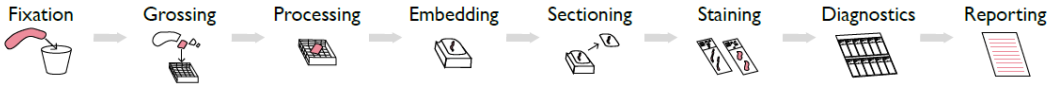


Figure 21. The histopathology laboratory workflow.

Currently, in the pathology laboratory in USA, more than 50,000 glass slides (range of 14,000 to 100,000 cases/specimens) are being stained per year and the scale is increasing. Since slides are glass based, it is very easy to get damaged and break and storing and archiving is not an easy task (Figure 22). Also, the information regards each slide is paper based and slides are reviewed manually by the microscope.

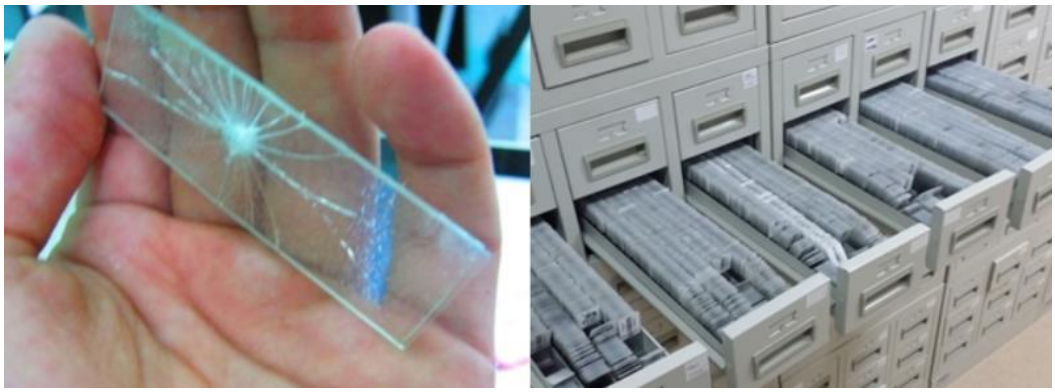


Figure 22. Storing glass slide is not an easy task.

Manual data entry is currently a bottle neck in the laboratory operations and carries the risk of data damage, loss and error. Besides the time consuming process, prone to errors, subject to human resource fluctuations and expensive to scale are the limitations of using manual data entry. Different approaches such as training the employees with the importance of the data and using software tool can reduce manual data entry errors.

Currently, using software tools to import data into computer systems is playing an important rule to help the archiving and tracking the information from each glass slide.

In the pathology laboratory, every tissue glass slide has labels which have been created by staining machines such as Benchmark Ultra or Autostainer (Figure 23).

Figure 23. Two example of staining machines used in pathology laboratories. Benchmark Ultra (left) and Autostainer (right).

Each label has meta data information related to that glass slide. Clone name, development, target name, detection method, tissue type, staining date, dilution factor, immunogen, AR field, autostainer and slide numbers are a number of metadata being mentioned on the slide label. Each glass slide with the label is scanned with the iScanHT scanner machine. Figure 24 demonstrates two examples of tissue glass slides with two different slide label platform.

Figure 24. Two examples of glass slide tissues with different labels platform. Ultra label (left) and autostainer label (right).

In order to reduce the manual work for slide tracking and archiving in the pathology laboratory and transferring all hard copy into a digital copy, a software tool with graphical user interface is designed to automatically extract metadata from tissue slide label images and store the information in the archive in digital form.

6.1.2. Methodology

Two types of tissue slides stained by Ultra and Autostainer platforms were scanned by iScanHT scanner and the label images are automatically extracted. Figure 25 demonstrates the interface window and the description of a different section of this interface is given as follow:

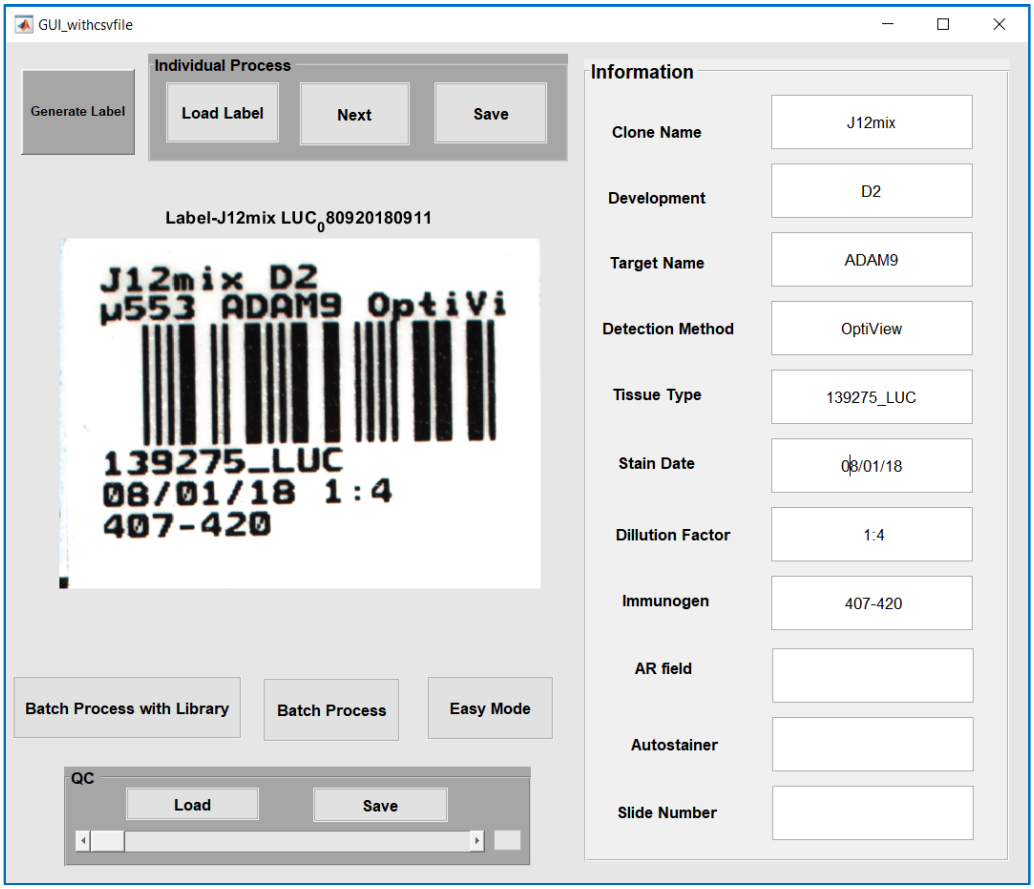


Figure 25. Graphical user interface window.

I. Generate Label

iScanHT scanner will scan the microscope slide and save the file as a tiff file which includes both tissue and label part of the glass slide. In order to process the labels, a generating algorithm is applied to create a .png file and extract the labels from the glass slide. By selecting the "Generate Label" the program asks the user to select the folder that contains different projects and start generating labels. A simple algorithm is used to select the labels region on the glass slide and save the ROI as a .png file. Since the workplace that we designed this software for is using a shared drive, loading and saving files from

the shared drive is taking a long time. With this key, user is able to put the program to work overnight and generate all the labels at one time.

II. Process

Different challenges with labels quality were faced during meta data information extraction process. Examples of challenges including, rotation of the label, stain leaked over the text part, scratch on the text part are shown in Figure 26.

Figure 26. Examples of the challenges with labels. Rotation, scratch on text part and stain over text.

Due to requests of the pathologist and challenges with the labels, different processing algorithms are designed to extract meta data information. Processing algorithms include: 1) individual process to be applied over labels individually and extract information one by one, 2) batch process to be applied over all the labels and extract information for all at once, 3) batch process with the library to be applied over a selection of the labels with having a set of similar meta information, and 4) is the easy mode which is only reading the information which is not similar among labels. The details of each process are being described in the following sections.

a) Individual Process

As mentioned before, the individual process is processing each label individually and one by one. By pressing the "Load Label", user is able to select the desired project folder containing labels and start the process. First, the program used a neural network based image classification algorithm to identify the label type and decide to choose the ultra or autostainer methods for extracting meta data information. Next, a set of pre image processing algorithms were applied to generate a better quality of the image and prepare image for text reading. Afterward our trained optical character recognition algorithm was applied to identify the texts in each label image. Based on the label type, each text will be assigned to a specific item for meta data information and the results will be shown in the information panel. Pressing the "Next" key allow users to go through each label individually and process the result one by one. Although using this process is time consuming but the accuracy of the result is high since the user is able to go through each label and check the result. Finally, the "Save" key allows users to export and save their extracted metadata into a csv file. This csv file later is used for uploading into a digital

pathology platform where the tissue slide images were hosted. Figure 27 shows the keys for individual process.

Figure 27. Individual process panel

b) Batch Process

This key processes all the images of a project at the same time and the results will be saved automatically into a csv file. By pressing the "Batch process" key, the interface ask user to select a folder to contain the labels and start processing all the labels at the same time. The process steps are similar to individual process, starting with a neural network to identify label type, using a set of –pre image processing to increase the label quality and then applying optical character recognition algorithms to extract text and assign each text item to a specific meta data information. The end of the process will be shown with a message box of "Operation Completed". Using batch process is much more efficient than individual process due to taking less time than individual process and availability to work overnight.

c) Batch process with library

In pathology labs, in the process of generating slide tissues, there will be glass slides with similar information such as tissue and target. The batch process with the library process is designed based on projects that have similar tissue or target information. The process is similar to batch processing with an additional step. By pressing the batch

process with a library key, a new window (library selection, Figure 28) is shown to the user to select the desired target and tissue. Pathologists prepared a list of the most common target and tissue items and these lists are stored in the program library. When user choose “Target” or “Tissue” key, the library related to it will be loaded to program and ask users what target or tissue name is desired (Figure 29).

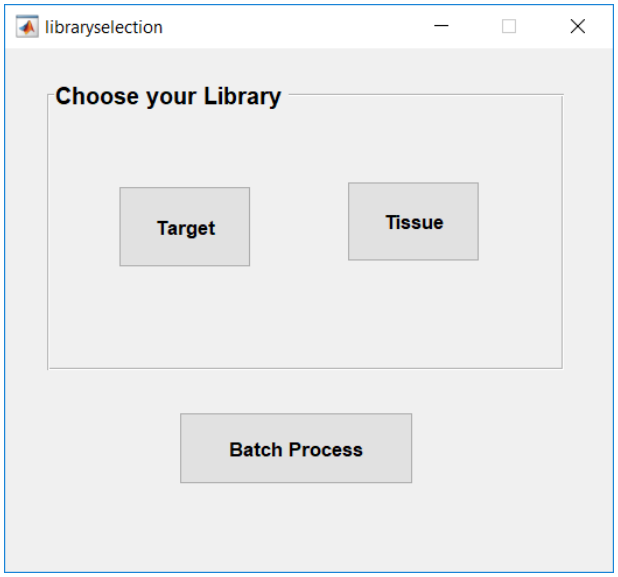
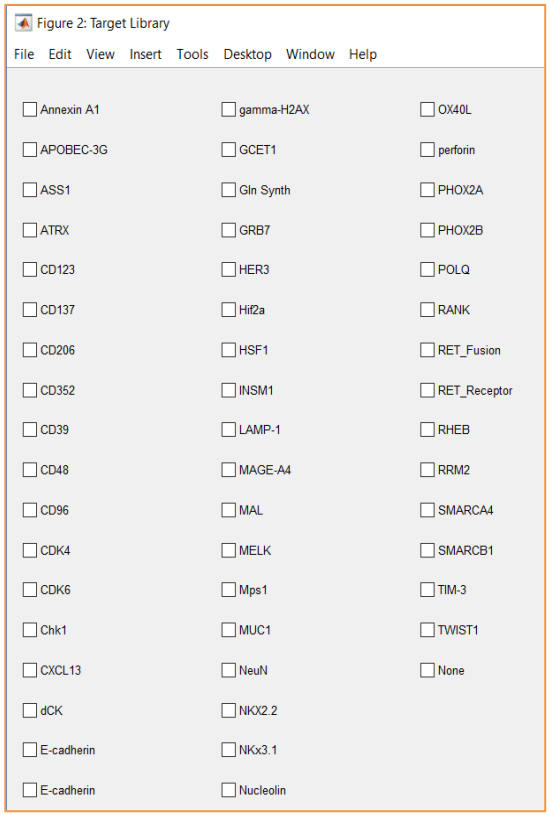


Figure 28. Library selection window for target and tissue.



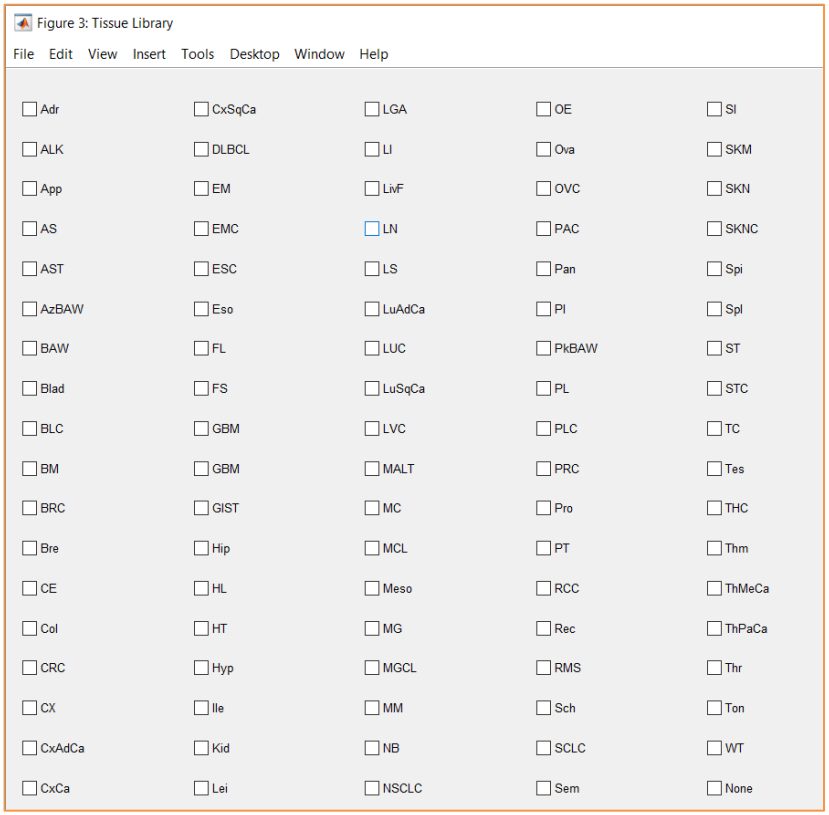


Figure 29: List of target and tissue library.

User is allowed to select as many target or tissue desired. This process helps the program to have a reference for target and tissue information and if there is scratch or stain over text, program is able to retrieve the information based on the references selected by the user from the library. As an example, Figure 30 demonstrates a label with difficulties in reading the tissue information. Due to the difficulties in reading the information, optical character recognition is not able to extract the correct information from the image for tissue information. Using the tissue library and selecting some references, program gives a score to the possibilities of each selected library. The reference with the highest score will be selected as the final result. The correct answer for tissue information for figure 10 is 158261_PL, the first result from OCR will be 158261_PI. If user select different references such as “PL”, “PLC” as tissue library, program will give scores to each reference based on the similarity with the result of the OCR. Since OCR got the tissue as “PI”, the reference of PL will receive higher scores more than “PLC” and the program will give “158261_PL” as the final result.

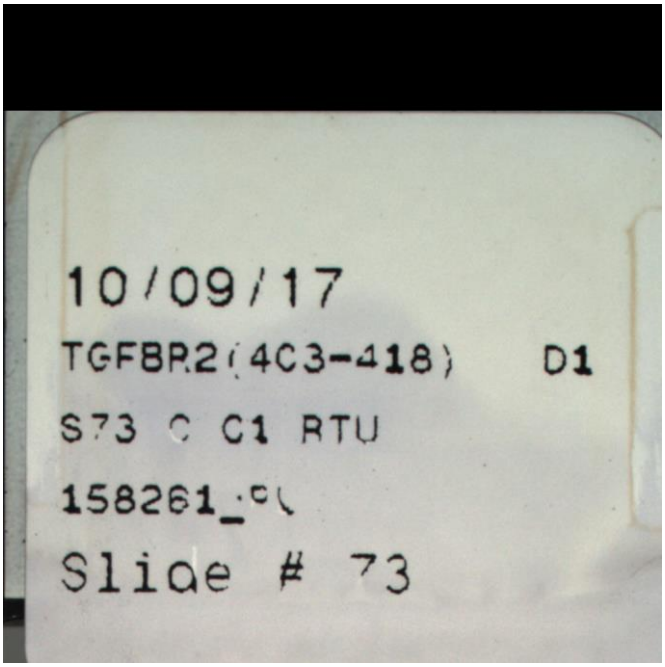


Figure 30. Example of missing text on label resulting difficulties in generating correct information.

d) Easy mode

Due to the requests of the pathologist, “Easy mode” process was designed to make the progress faster. For this process, user is having a specific project containing different labels with similar information in different sections and only one meta data is different. As an example, there will be projects defined to have similar label types, development, detection method, target, tissue, stain date, dilution facto and immunogen and only clone name is different. With this information entered into the program, program will read only the information that is not given and less time will be put over processing. By selecting the “Easy mode”, the name selection window will show up and allow user to decide what are the similarities among the labels. Pathologists provided a list of all common meta data information for each item which can be updated in the future. By selecting each key from “Select Information” panel, a list of all available information will

be shown and allow user to select one item which is going to be similar among all labels. Figure 31 shows the “Select Information” panel and an example of one of the information selected.

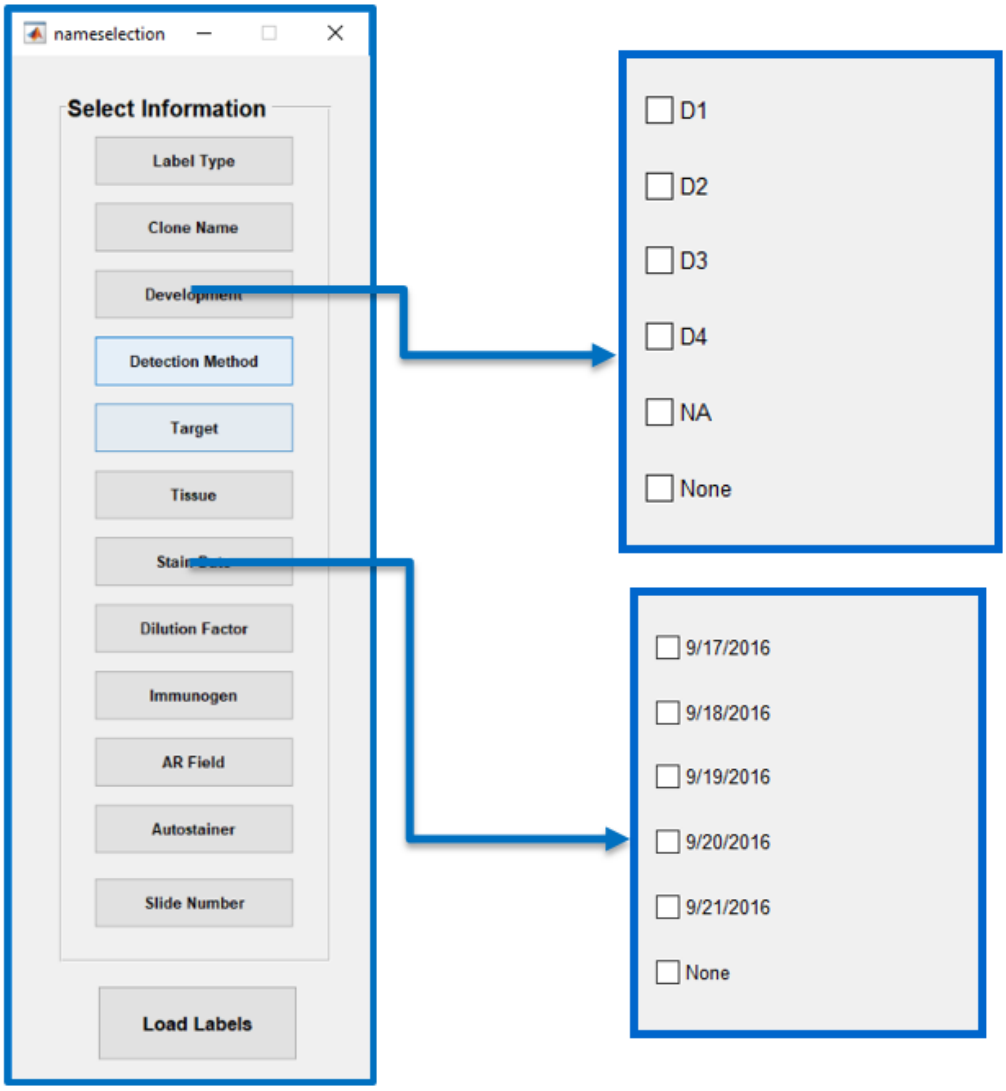


Figure 31. By pressing “Easy mo*de” key, a window with “select* formation” panel will show (left) and by selecting each item, a list of all available item will be shown (right).

III. Quality Control:

In order to allow user to easily check the result of the interface and the extraction of the meta data, a quality control section is added to the program. This panel allows user to visually see the label processed and show the result of each meta data. Figure 32 demonstrates the quality control panel with the keys implemented. “Load” key will load all the labels in a project and the slide bar allow user to go through each label and enter any correction needed in the main information panel. In the end, user is able to save the updated information and update the previously generated csv file.

Figure 32. Quality control panel with the keys and slide bar.

6.1.3 Conclusion

This tool was tested over 277 tissue slide labels with different challenges including label damage, rotation, stain over texts, etc. Label type identification achieved 100% accuracy and all labels correctly identified as Ultra or Autostainer label. The individual screening method took one hour as the user visually examined the results during the process through the tool’s user interface and received the higher accuracy. Batch Processing is more efficient since it only took one minute to process the dataset with an error rate of 15%. Further improvement can be achieved by developing a dedicated quality control interface used at the end of batch processing.

Chapter 7: Conclusion and Future Work

7.1 Summary

To improve efficacy of disease diagnosis and treatment, a recent precision medicine initiative calls for developing a new disease treatment strategy that takes individual variability into account [97]. In response to this initiative, the researchers in medical imaging informatics field hypothesized that developing and applying computer-aided detection and diagnosis (CAD) schemes can play an important and unique role to help establish the new precision or personalized medicine paradigm. Besides the conventional CAD schemes that focus on detecting suspicious lesions and serve as "a second reader" to help radiologists detect more cancers in cancer screening environment, new and novel CAD schemes have been developed in recent years, which have demonstrated their capability and advantages to much broad clinical applications, such as predicting short-term risk of cancer development, classifying between malignant and benign lesions, determining disease severity and types, predicting disease prognosis, assessing efficacy of treatment and clinical outcome (i.e., disease-free or progression-free survival).

In the previous chapters of this dissertation, I present and report several new CAD schemes, which were developed and evaluated in my PhD studies. In Chapters 2 and 3 I present and discuss new CAD schemes of breast MRI acquired pre- and post-neoadjuvant chemotherapy of breast cancer patients to predict radiographic complete response and pathologic complete response of breast cancer patients to the therapies. In Chapter 4 I present and discuss a new interactive CAD scheme to detect, segment and quantify severity of aneurysmal subarachnoid hemorrhage (aSAH) of stroke patients using brain

CT images. In Chapter 5 I present and discuss another CAD schemes of digital pathology images using both histogram-based features computed from low resolution images and convolution neural network (CNN) generated features computed from the high resolution images to identify the ROIs in which tissues or cells are responsive to the specific antibody stains. In Chapter 6 I present and discuss a unique and efficient CAD scheme with optical character recognition algorithm to identify the label type between ultra and autostainer of pathology images to facilitate the efficiency of digital pathology image storage.

The effort and contribution of my PhD research work include (1) developing and testing new image processing algorithms to segment regions of interest (i.e., breast areas in MRI, blood or aSAH regions in brain CT images, and antibody-responsive regions or cells in digital pathology images), (2) selecting optimal image features from the initially large feature pool, (3) training and cross-validating machine learning models or classifiers to minimize bias and overfitting, and (4) developing and implementing interactive CAD schemes with easy-to-use graphic user interface to increase the flexibility of the using CAD scheme and confidence of the clinicians to accept CAD-generated results. Through these studies, we published following papers in which I am a co-author during the period of my PhD studies.

A. Journal Papers

1. **Aghaei F**, Tan M, Hollingsworth AB, Qian W, Liu H, Zheng B, Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy, *Medical Physics* 2015; 42:6520-6528.

2. **Aghaei F**, Tan M, Hollingsworth AB, Zheng B, Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy, *Journal of Magnetic Resonance Imaging*, 2016; 44:1099-1106.
3. Tan M, **Aghaei F,** Wang Y, Zheng B, Developing a new case based computer-aided detection scheme and an adaptive cueing method to improve performance in detecting mammographic lesions, *Physics in Medicine and Biology* 2017; 62:358-376.
4. Wang Y, **Aghaei F,** Zarafshani A, Qiu Y, Qian W, Zheng B, Computer-aided classification of mammographic masses using visually sensitive image features, *Journal of X-ray Science and technology* 2017; 25:171-186.
5. Yan S, Wang Y, **Aghaei F,** Qiu Y, Zheng B, Applying a new bilateral mammographic density segmentation method to improve accuracy of breast cancer risk prediction, *International Journal of Computer Assisted Radiology and Surgery*, 2017; 12; 1819-1828.
6. Yan S, Wang Y, **Aghaei F,** Qiu Y, Zheng B, Improving performance of breast cancer risk prediction by incorporating optical density image feature analysis: An assessment, *Academic Radiology* 2017; S1076-6332(17)30364-1.
7. Danala G, Patel B, **Aghaei F,** Heidari M, Li J, Wu T, Zheng B, Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms, *Annals of Biomedical Engineering*, 2018; 46:1419-1431.
8. Santucci J, Ross S, Greemert J, **Aghaei F,** Ford L, Hollabaugh K, Conwell B, Wu D, Zheng B, Bohnstedt B, Ray B, Radiological estimation of intracranial blood volume and occurrence of hydrocephalus determines stress-induced

hyperglycemia after aneurysmal subarachnoid hemorrhage, *Translational Stroke Research* 2019; 10:327-337.

9. Ray B, Ross SR, Danala G, Aghaei F, Nouh CD, Ford L, Hollabaugh KM, Karfonta BN, Santucci JA, Cornwell BO, Bohnstedt BN, Zheng B, Dale GL, Prodan CI, Systemic response of coated-platelet and peripheral blood in flammatory cell indices after aneurysmal subarachnoid hemorrhage and long-term clinical outcome, *Journal of Critical Care,* 2019; Doi: 10.1016/jcc.2019.03.003.

B. Conference Proceeding Papers

1. Tan M, **Aghaei F,** Wang Y, Qian W, Zheng B, Improving the performance of lesion-based computer-aided detection schemes of breast masses using a case-based adaptive cueing method, *Proc SPIE* 2016; 97851V.
2. **Aghaei F,** Tan M, Hollingsworth AB, Zheng B, Cheng S, Computer-aided global breast MR image feature analysis for prediction of tumor response to chemotherapy: performance assessment, *Proc SPIE* 2016; 978522.
3. Tan M, **Aghaei F,** Hollingsworth AB, Stough RG, Liu H, Zheng B, Increasing cancer detection yield of breast MRI using a new CAD scheme of mammograms, *Proc SPIE* 2016; 97850R.
4. Gaffney KP, **Aghaei F,** Battiste J, Zheng B, Automated detection and quantification of residual brain tumor using an interactive computer-aided detection scheme, *Proc SPIE* 2017; 101342I.
5. **Aghaei F,** Mirniaharikandehei S, Hollingsworth AB, Wang Y, Qiu Y, Liu H, Zheng B, Exploring a new bilateral focal density asymmetry based image marker to predict breast cancer risk, *Proc SPIE* 2017; 101361P.

6. Wang Y, **Aghaei F,** Tan M, Qiu Y, Liu H, Zheng B, Developing a visual sensitive image features based CAD scheme to assist classification of mammographic masses, *Proc SPIE* 2017; 101361M.
7. **Aghaei F,** Ross SR, Wang Y, Wu DH, Cornwell BO, Ray B, Zheng B, Implementation of a computer-aided detection tool for quantification of intracranial radiologic markers on brain CT images, *Proc SPIE* 2017; 1013805.
8. Mirniaharikandehei M, Patil O, **Aghaei F,** Wang Y, Zheng B, Exploring a new quantitative image marker to assess benefit of chemotherapy to ovarian cancer patients, *Proc SPIE* 2017; 101380I.
9. Zarafshani A, Dhurjaty S, Mirniaharikandehei S, **Aghaei F,** Xiang L, Zheng B, Developing a unique portable device to non-invasively detect ioelectrochemical characteristics of human tissues, *Proc SPIE* 2018; 10573-207.
10. Danala G, **Aghaei F,** Heidari M, Wu T, Patel B, Zheng B, Computer-aided classification of breast masses using contras-enhanced digital mammograms, *Proc SPIE* 2018; 105752K.
11. **Aghaei F,** Danala G, Wang Y, Zarafshani A, Qian W, Liu H, Zheng B, Association between mammogram density and background parenchymal enhancement of breast MRI, *Proc SPIE* 2018; 105752O.
12. Mirniaharikandehei S, Zarafshani A, Heidari M, Wang Y, **Aghaei F,** Zheng B, Applying a CAD-generated imaging marker to assess short-term breast cancer risk, *Proc SPIE* 2018; 105753F.

13. **Aghaei F,** Danala G, Hollingsworth AB, Stoug RG, Pearce M, Liu H, Zheng B, Applying a new mammographic imaging marker to predict breast cancer risk, *Proc SPIE* 2018; 105753N.
14. **Aghaei F,** Mirniaharikandehei S, Hollingsworth AB, Stoug RG, Pearce M, Liu H, Zheng B, Association between background parenchymal enhancement of breast MRI and BIRADS rating change in the subsequent screening, *Proc SPIE* 2018; 105790R.
15. **Aghaei F,** Hollingsworth AB, Mirniaharikandehei S, Wang Y, Liu H, Zheng B, Developing a new quantitative imaging marker to predict pathological complete response to neoadjuvant chemotherapy, *Proc SPIE* 2019; 109502O.
16. Zarafshani A, Wang Y, Mirniaharikandehei S, Heidari M, **Aghaei F,** Wang S, Xiang L, Zheng B, Design, fabrication and evaluation of non-imaging, label-free pre-screening tool using quantified bio-electrical tissue profile, *Proc SPIE* 2019; 1095304.
17. Danala G, Heidari M, **Aghaei F,** Ray B, Zheng B, Developing a computer-aided image analysis and visualization tool to predict region-specific brain tissue "at risk" for developing acute ischemic stroke, *Proc SPIE* 2019; 109530M.
18. **Aghaei F**, Nie Y, Computer Aided Antibody Screening for IHC Assay Development, *Proc SPIE* 2019; 109560C.

In summary, besides the publications, the most significance or contribution of my research work is that several CAD schemes and GUI tools discussed in this dissertation have been implemented and used by the clinical researchers in OUHSC and biomedical engineering researchers in Roche company. The promising study or test results have been

reported in these independent studies. Thus, my research work demonstrates the feasibility of developing and applying new novel CAD schemes to extract quantitative image markers and build machine learning models for variety of medical image based disease detection, diagnosis and prognosis assessment. The interactive CAD schemes with GUI can provide clinicians the decision-making support tools to help them more accurately diagnosis diseases and make optimal treatment decisions for the patients.

7.2 Future works

Although great research efforts have been made in developing and testing new CAD schemes for many different clinical application purposes, many technical challenges remain to develop the CAD schemes with high scientific rigor (or robustness) due to several factors, which include (1) difficulty and errors in segmentation of disease or other related regions of interest, (2) potential bias in feature computation and selection, (3) the limited size of image datasets to train and cross-validate machine learning models, and (4) the gap between computer vision and human vision. Thus, many development and optimization issues still need to be better investigated and solved before any CAD schemes can be optimally implemented and used by the clinicians in the clinical practice. Based on the knowledge and hand-on experience that I have leant from my previous studies, I will continue investigating these issues and exploring novel approaches trying to make new progress or contribution to help solve these issues in my future work.

For example, I will continue focus on developing and implementing new CAD schemes of digital pathology images. Every year, an average of approximately 50,000

cytology specimens are produced.. Pathology images can be equal to information sent from a satellite in space over street map regard size which means much information exists in one whole slide digital pathology image. The combination of image analysis and machine learning over digital pathology images can be used to automatically segment cells and detect normal and tumor cells. The research scientist and pathologist will continue to be instrumental in both the use and operation of image analysis workflows, which will continue to evolve preclinical and clinical biomarker research. Also, taking advantage of new research topics such as deep learning on pathology images can help in improving the detection and analysis of digital pathology images. In conclusion, my next short-term goal is to work on detection and segmentation of tumor cells depicted in digital pathology images scanned by two different scanners available in pathology laboratories.

References

1. L. B. Lusted, “Logical analysis in roentgen diagnosis - Memorial fund lecture,” Radiology **74**, 178–193 (1960).
2. W. J. Tuddenham, “Visual search, image organization, and reader error in roentgen diagnosis – Studies of the psychophysiology of roentgen image perception -Memorial fund lecture,” Radiology **78**, 694–704 (1962).
3. H. L. Kundel and G. Revesz, “Lesion conspicuity, structured noise, and film reader error,” Am. J. Roentgenol. **126**, 1233–1238 (1976).
4. K. S. Berbaum, E. A. Franken, D. D. Dorfman, S. A. Rooholamini, M. H. Kathol, T. J. Barloon, F. M. Behlke, Y. Sato, C. H. Lu, G. Y. Elkhoury, F. W. Flickinger, and W. J. Montgomery, “Satisfaction of Search in diagnostic-radiology,” Invest. Radiol. **25**, 133–140 (1990).
5. D. L. Renfrew, E. A. Franken, K. S. Berbaum, F. H. Weigelt, and M. M. Abuyousef, “Error in radiology – Classification and lessons in 182 cases presented at a problem case conference,” Radiology **183**, 145-150 (1992).
6. J. Dheeba, N. Albert Singh, S. Tamil Selvi, “Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, “J. Biomedical Informatics **49**, 45-52 (2014).
7. X. Sun, W. Qian, D. Song, “Three-class classification in computer-aided diagnosis of breast cancer by support vector machine,” Proc. SPIE 5370 (2004).
8. S. E. Song, B. K. Seo, K. R. Cho, O. H. Woo, G. S. Son, C. Kim, S. B. Cho, S. S. Kwon, “Computer-aided detection (CAD) system for breast MRI in

assessment of local tumor extent, nodal status, and multifocality of invasive breast cancers: preliminary study" Cancer Imaging 15:1(2015).

9. Y. Yuan, M. L. Giger, H. Li, N. Bhooshan, C. A. Sennett, "Multimodality computer-aided breast cancer diagnosis with FFDM and DCE-MRI," *Acad Radiol*. **17(9)**,1158-1167 (2010).

10. M. D. Schnall, "Breast MR imaging," Radiol Clin N AM **41**, 43-50 (2003).

11. M. T. Mandelson, N. Oestreicher, P. L. Porter, D. White, C. A. Finder, S. H. Taplin, E. White, "Breast Density as a Predictor of Mammographic Detection: Comparison of Interval- and Screen-Detected Cancers," J Natl C Inst 92, 1081-1087 (2000).

12. U. Fischer, L. Kopka, E. Grabbe, "Breast carcinoma: effect of preoperative contrast-enhanced MR imaging on the therapeutic approach," Radiology 213, 881-888 (1999).

13. J. J. Fenton, J. Wheeler, P. A. Carney, G. Cutter, C. D'Orsi, E. A. Sickles, J. Fosse, L. Abraham, S. H Taplin, W. Barlow, R. E. Hendrick, J. G. Elmore, "Reality check: perceived versus actual performance of community mammographers," AJR Am J Rotengenol **187**, 42–46 (2006).

14. L. Berlin, F. M. Hall, „More mammography muddle: emotions, politics, science, costs and polarization," Radiology **255**, 311–316 (2010).

15. D. Saslow, C. Boetes, W. Burke S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, R. A. Smith, E. Warner, M. Yaffe, K. S. Andrews, C. A. Russell, "American Cancer Society guidelines for breast

screening with MRI as an adjunct to mammography," CA Cancer J Clin **57** ,75–89 (2007).

16. J. G. M. Klijn, "Early diagnosis of hereditary breast cancer by magnetic resonance imaging: What is realistic?," J Clin Oncology **28**, 1441–1445 (2010).

17. L. J. Graham, M. P. Shupe, E. J. Schneble, F. L. Flynt, M. N. Clemenshaw, A. D. Kirkpatrick, C. Gallagher, A. Nissan, L. Henry, A. Stojadinovic, G. E. Peoples, N. M. Shumway, "Current approaches and challenges in monitoring treatment response in breast cancer," J Cancer **5**, 58–68 (2014).

18. R. Bekkerman, M. Bilenko, J. Langford, "Scaling up Machine Learning: Parallel and Distributed Approaches," Cambridge, Cambridge University Press., (2011).

19. T. M. Mitchell, "Machine Learning" McGraw Hill, Inc., New York, NY (1997).

20. V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, New York, Inc., New York, NY (1995).

21. X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, D. Steinberg, "Top 10 Algorithms in Data Mining", Knowledge and Information Systems (14:1), 1-37 (2007).

22. S. B. Kotsiantis, « Supervised Machine Learning: A Review of Classification Techniques," Informatica **31**, 249-268 (2007).

23. D. Wettschereck, D. W. Aha, T. Mohri, “A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms,” Artificial Intelligence Review 10, 1-37, (1997).

24. C. Bernhard Scholkopf, J. C. Burges, A. J. Smola, “Advances in Kernel Methods: support vector learning,” MIT Press, Cambridge, MA (1999).

25. C. Neocleous, C. Schizas, “Artificial Neural Network Learning: A Comparative Review,” LNAI 2308, 300-S313, Springer (2002).

26. J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Networks, **61**, 85-117 (2015).

27. A. S. Weigend, D. E. Rumelhart, B. A. Huberman, “Generalization by weight elimination with application to forecasting,” Proceedings of the conference on Advances in neural information processing systems **3**, 875-882, Denver, Colorado, USA (1991).

28. R. P. Lippmann, J. Moody, D. S. Touretzky, “Advances in Neural Information Processing Systems 3,” San Mateo, CA, Morgan Kaufmann Publishers Inc. (1990)

29. M. N. H. Siddique, M. O. Tokhi, “Training Neural Networks: Backpropagation vs. Genetic Algorithms,” IEEE International Joint Conference on Neural Networks, **4**, 2673-2678 (2001).

30. S. K. Murthy, “Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey,” Data Mining and Knowledge Discovery **2**, 345-389 (1998).

31. J. R. Quinlan, “C4.5: Programs for machine learning,” Morgan Kaufmann Publishers Inc., San Francisco, CA (1993).

32. J. R. Quinlan, “Discovering rules by induction from large collections of examples,” In Expert Systems in the Microelectronic Age; Michie, D. Ed, Edinburgh University Press: Edinburgh, UK, 168-201 (1979).

33. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, “Classification and Regression Trees,” Wadsforthand Brooks Publisher Inc., Monterey, CA, (1984).

34. E. Hunt, J. Martin, P. Stone, “Experiments in Induction,” New York, Academic Press, New York, NY (1966).

35. I. Kononenko, “Estimating attributes: Analysis and extensions of RELIEF,” F. Bergadano, L. De Raedt, eds., Proceeding European Conference on *Machine Learning*, Springer Verlag, Berlin, 171-182 (1994).

36. T. Elomaa, “The Biases of Decision Tree Pruning Strategies,” Proceedings of the Third International Symposium on Advances in Intelligent Data Analysis, 63-74 (1997).

37. A. Vlachos, “Active learning with support vector machines,” Computer sciences technical report, University of Edinburgh (2004).

38. B. Settles, “Active learning literature survey,” Computer Sciences Technical Report, University of Wisconsin–Madison (2009).

39. D. Yu, B. Varadarajan, L. Deng, A. Acero, “Active learning and semi-supervised learning for speech recognition: A unified framework using the

global entropy reduction maximization criterion," Journal Computer Speech and Language **24**, 433-444 (2010).

40. D. Sculley, "Online active learning methods for fast label-efficient spam filtering," In Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. (2007).

41. F. Sardanelli, F. Podo, "Breast MR imaging in women at high-risk of breast cancer. Is something changing in early breast cancer detection?" Eur Radiol **17**, 873-887 (2007).

42. W. A. Berg, Z. Zhang, D. Lehrer, et al., "Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk," JAMA **307**, 1394-1404 (2012).

43. N. Bhooshan, M. L. Giger, S. A. Jansen, H. Li, L. Lan, "Newstead GM. Cancerous breast lesions on dynamic contrast-enhanced MR images: Computerized characterization for image-based prognostic markers," Radiology **254**, 680-690 (2010).

44. B. Fisher, J. Bryant, N. Wolmark, "Effect of preoperative chemotherapy on the outcome of women with operable breast cancer," J Clin Oncol. **16**, 2672-2685 (1998).

45. J. A. Van der Hage, C. J. Van de Velde, J. P. Julien, "Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer Trial 10902," J Clin Oncol. **19(22)**, 4224-4237 (2001).

46. D. Mauri, N. Pavlidis, J. P. Ioannidis, “Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis,” J of the National Cancer Inst. **97(3)**, 188-194 (2005).

47. M. B. Lobbes, R. Prevos, M. Smidt, V. C. Tian-Heijnen, et al., “The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: a systematic review,” Insights Imaging **4**, 163-175 (2013).

48. M. L. Marinovich, N. Houssami, P. Macaskill, et al, “Meta-analysis of magnetic resonance imaging in detecting residual breast cancer after neoadjuvant therapy,” J Natl Cancer Inst **105**, 321-333 (2013).

49. E. A. Eisenhauer, P. Therasse, J. Bogaerts, et al, “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1),” Eur J Cancer. **45**, 228-247 (2009).

50. R. Rouzier, C. M. Perou, W. F. Symmans, et al, “Breast cancer molecular subtypes respond differently to preoperative chemotherapy,” Clin Cancer Res. **11**, 5678-5685 (2005).

51. F. Aghaei, M. Tan, A. B. Hollingsworth, W. Qian, H. Liu, B. Zheng, “Computer-aided Breast MR Image Feature Analysis for Prediction of Tumor Response to Chemotherapy,” Med Phys. **42**, 6520-6528 (2015).

52. V. King, J. D. Brooks, J. Bernstein, “Background parenchymal enhancement at breast MR imaging and breast cancer risk,” Radiology **260**, 50-60 (2011).

53. Q. Yang, L. Li, J. Zhang, G. Shao, B. Zheng, "A computerized global MR image feature analysis scheme to assist diagnosis of breast cancer: a preliminary assessment," Eur J Radiol **83**, 1086-1091 (2014).

54. W. Chen, M. L. Giger, U. Bick, G. M. Newstead, "Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI," Med. Phys. **33**, 2878-2887 (2006).

55. I. Witten, E. Frank, M. A. Hall, "Data mining: practical machine learning tools and techniques, 3rd ed," Elsevier, http://www.cs.waikato.ac.nz/ml/weka/.

56. R. Kohavi, G. H. John, "Wrappers for feature subset selection," Artificial Intelligence **97**, 273-324 (1997).

57. Q. Li, K. Doi, "Reduction of bias and variance of evaluation of computer-aided diagnostic schemes," Medical Physics **33**, 868-875 (2006).

58. S. C. Anger, M. A. Rosen, S. Englander, et al, "Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: A feasibility study," Radiology **272**, 91-99 (2014).

59. D. V. Schacht, K. Drukker, I. Pak, H. Abe, M. L. Giger, "Using quantitative image analysis to classify axillary lymph nodes on breast MRI: a new application for the Z 0011" Era, Eur J Radiol **84**, 392-397 (2015).

60. B. Zheng, J. H. Sumkin, M. L. Zuley, X. Wang, A. H. Klym, D. Gur, "Bilateral mammographic density asymmetry and breast cancer risk: A preliminary assessment," European Journal of Radiology **81**,3222-3228 (2012).

61. P. Lambin, E. Rios-Velazquez, R. Leijenaar, et al, “Radiomics: extracting more information from medical images using advanced feature analysis,” Euro J Cancer **48**, 441-446 (2012).

62. H. Aerts, et al. Decoding tumor phenotype by noninvasive imaging using a quantitative Radiomics approach. Nat. Commun. **5**, 4006 (2014).

63. M. Teshorme, K. Hunt, “Neoadjuvant therapy in treatment of breast cancer,” Surg Oncol Clin N Am. **23**, 505-523 (2014).

64. J. S. Vaidya, S. Massarut, et al., “Rethinking neoadjuvant chemotherapy for breast cancer,” BMJ.; 360:j5913 (2018).

65. J. R. Gralow, H. J. Burstein, W. Wood, G. N. Hortobagyi, L. Gianni, G. von Minckwitz, A. U. Buzdar, I. E. Smith, W. F. Symmans, B. Singh, E. P. Winer, “Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease,” J Clin Oncol **26**, 814–19 (2008).

66. P. Cortazar, L. Zhang, M. Untch, et al. “Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis,” Lancet, **384**, 164-172 (2014).

67. W. F. Symmans, C. Wei, R. Gould, et al., “Long-term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype,” J Clin Oncol **35**,1049-1060 (2017).

68. T. Fukuda, R. Horii, N. Gomi, et al., “Accuracy of magnetic resonance imaging for predicting pathological complete response of breast cancer after neoadjuvant chemotherapy: association with breast cancer subtype,” Springerplus, 5:152 (2016).

69. M. Kaufmann, G. N. Hortobagyi, A. Goldhirsch, S. Scholl, A. Makris, P. Valagussa, J. U. Blohmer, W. Eiermann, R. Jackesz, W. Jonat, A. Lebeau, S. Loibl, W. Miller, S. Seeber, V. Semiglazov, R. Smith, R. Souchon, V. Stearns, M. Untch, G. von Minckwitz, "Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: An Update," J Clin Oncol **24**, 1940-1949 (2006).

70. M. Basik, J. P. Costantino, J. F. De Los Santos, et al. "Oncology BR005: phase II trial assessing accuracy of tumor bed biopsies (Bx) in predicting pathologic response in patients (Pts) with clinical/radiological complete response (CR) after neoadjuvant chemotherapy (NCT) in order to explore the feasibility of breast-conserving treatment (BCT) without surgery,". *J Clin Oncol*. **36** (15_suppl), TPS604 (2018).

71. N. Wolmark, J. Wang, E. Mamounas, J. Bryant, B. Fisher, "Preoperative chemotherapy in patients with operable breast cancer: nine-year results from National Surgical Adjuvant Breast and Bowel Project B-18," *J Natl Cancer Inst Monogr* **30**, 96–102 (2001).

72. G. von Minckwitz, M. Untch, J. U. Blohmer, S. D. Costa, H. Eidtmann, P. A. Fasching, B. Gerber, W. Eiermann, J. Hilfrich, J. Huober, C. Jackisch, M. Kaufmann, G. E. Konecny, C. Denkert, V. Nekliudova, K. Mehta, S. Loibl, "Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes," *J Clinical Oncol* **30**, 1796-804 (2012).

73. J. R. Scheel, E. Kim, S. C. Partridge, C. D. Lehman, M. A. Rosen, W. K. Bernreuter, E. D. Pisano, H. S. Marques, E. A. Morris, P. T. Weatherall, S. M. Polin, G. M. Newstead, L. J. Esserman, M. D. Schnall, N. M. Hylton, "MRI, Clinical Examination, and Mammography for Preoperative Assessment of Residual Disease and Pathologic Complete Response After Neoadjuvant Chemotherapy for Breast Cancer: ACRIN 6657 Trial, "American Journal of Roentgenology, 210(6), 1376-1385 (2018).

74. F. Aghaei, M. Tan, A. B. Hollingsworth, B. Zheng, "Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy," Journal of Magnetic Resonance Imaging **44**, 1099-1106 (2016).

75. R. M. Haralick, K. Shanmugam., L. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man. Cybern **6**, 610-621 (1973).

76. M. M. Galloway, "Texture analysis using grey level run lengths," Comput Graphics Image Process **4**, 172-179 (1975).

77. A. Pennisi, T. Kieber-Emmons, I. Makhoul, L. Hutchins, "Relevance of Pathological Complete Response after Neoadjuvant Therapy for Breast Cancer" *Breast Cancer: Basic and Clinical Research* **10**, 103–106 (2016). doi:10.4137/BCBCR.S33163.

78. M. S. Aapro, "Neoadjuvant therapy in breast cancer: can we define its role?" Oncologist **6**, Suppl. 3, 36–9 (2001).

79. S. D. Heys, A. W. Hutcheon, T. K. Sarkar, K. N. Ogston, I. D. Miller, S. Payne, I. Smith, L. G. Walker, O. Eremin, Aberdeen Breast Group,

"Neoadjuvant docetaxel in breast cancer: 3-year survival results from the Aberdeen trial," Clin Breast Cancer **3**, Suppl. 2, 69-74 (2002).

80. J. B. Bederson, E. S. Connolly, H. H. Batjer, et al., "Guidelines for the management of aneurysmal subarachnoid hemorrhage: a statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association," Stroke, 40(3), 994-1025 (2009).

81. B. M. Kissela, L. Sauerbeck, D. Woo, et al., "Subarachnoid hemorrhage: a preventable disease with a heritable component," Stroke, 33(5), 1321-1326 (2002).

82. W. Brinjikji, D. F. Kallmes, G. Lanzino, H. J. Cloft, "Hospitalization costs for endovascular and surgical treatment of ruptured aneurysms in the United States are substantially higher than Medicare payments," American journal of neuroradiology, 33(6), 1037-1040 (2012).

83. T. N. Taylor, P. H. Davis, J. C. Torner, et al., "Lifetime cost of stroke in the United States," Stroke, 27(9), 1459-1466 (1996).

84. J. Ryu, T. Kim, C. Jeong, et al., "Dedicated mobile volumetric cone-beam computed tomography for brain imaging: a phantom study," Journal of X-ray Science and Technology, 23(4), 473-480 (2015).

85. J. Li, X. Tang, B. Wang, et al., "Comparison between dual arc VMAT and 7F-IMRT in the projection of hippocampus for patients during whole brain radiotherapy," Journal of X-ray Science and Technology, 24(3), 457-466 (2016).

86. S. B. Ko, H. A. Choi, A. M. Carpenter, et al., “Quantitative analysis of hemorrhage volume for predicting delayed cerebral ischemia after subarachnoid hemorrhage,” Stroke, 42(3), 669-674 (2011).

87. L. Jimenez-Roldan, J. F. Alen, P. A. Gomez, et al., “Volumetric analysis of subarachnoid hemorrhage: assessment of the reliability of two computerized methods and their comparison with other radiographic scales,” Journal of neurosurgery, 118(1), 84-93 (2013).

88. X. Wang, L. Li, W. Liu, W. Xu, D. Lederman, B. Zheng, “An interactive system for computer-aided diagnosis of breast masses,” Journal of Digital Imaging, 25(5), 570-579 (2012).

89. B. Zheng, “Applying quantitative medical image processing and feature analysis to assess near-term breast cancer risk,” Optical Instruments, 38(3), 189-199 (2016).

90. Y. Wang, F. Aghaei, A. Zarafshani, et al., “Computer-aided classification of mammographic masses using visually sensitive image features,” Journal of X-ray Science and technology, DOI:10.3233/XST-16212 (2016).

91. M. Tan, Z. Li, Y. Qiu, et al., “A new approach to evaluate drug treatment response of ovarian cancer patients based on deformable image registration,” IEEE Transactions on Medical Imaging, 35(1), 316-325 (2016)

92. R. Raj, J. Siironen, M. B. Skrifvars, et al., “Predicting outcome in traumatic brain injury: development of a novel computerized tomography classification system (Helsinki computerized tomography score),” Neurosurgery, 75(6), 632-646 (2014).

93. A. I. Maas, C. W. Hukkelhoven, L. F. Marshall, E. W. Steyerberg, "Prediction of outcome in traumatic brain injury with computed tomographic characteristics: a comparison between the computed tomographic classification and combinations of computed tomographic predictors," Neurosurgery, 57(6), 1173-1182 (2005).

94. J. A. Santucci, S. R. Ross, J. C. Greemert, F. Aghaei, L. Ford, K. M. Hollabaugh, B. O. Conwell, D. Wu, B. Zheng**,** B. N. Bohnstedt, B. Ray, "Radiological estimation of intracranial blood volume and occurrence of hydrocephalus determines stress-induced hyperglycemia after aneurysmal subarachnoid hemorrhage," *Translational Stroke Research*; 10:327-337 (2019).

95. B. Ray, S. R. Ross, G. Danala, F. Aghaei, C. D. Nouh, L. Ford, K. M. Hollabaugh, B. N. Karfonta, J. A. Santucci, B. O. Cornwell, B. N. Bohnstedt, B. Zheng**,** G. L. Dale, C. I. Prodan, "Systemic response of coated-platelet and peripheral blood in flammatory cell indices after aneurysmal subarachnoid hemorrhage and long-term clinical outcome," *Journal of Critical Care (*2019); Doi: 10.1016/jcc.2019.03.003.

96. Y, Nie, C. Roessler, E. Andersson, O. Grimm, "Color deconvolution method with DAB scatter correction for bright field image analysis", Proceedings of SPIE Vol. 10581, 105810K, (2018).

97. F.S. Collins, H. Varmus, "A new initiative on precision medicine," The New England Journal of Medicine, 372(9), 793-795 (2015).