UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPING AND APPLYING CAD-GENERATED IMAGE MARKERS TO ASSIST DISEASE DIAGNOSIS AND PROGNOSIS PREDICTION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

GOPICHANDH DANALA

Norman, Oklahoma

2022

DEVELOPING AND APPLYING CAD-GENERATED IMAGE MARKERS TO ASSIST DISEASE DIAGNOSIS AND PROGNOSIS PREDICTION

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Bin Zheng, Chair

Dr. Le Gruenwald

Dr. Hong Liu

Dr. Samuel Cheng

Dedicated to my parents and family.

For their endless love, support, and encouragement.

Table of Contents

List of Figures

List of Tables

Abstract

Developing computer-aided detection and/or diagnosis (CAD) schemes has been an active research topic in medical imaging informatics (MII) with promising results in assisting clinicians in making better diagnostic and/or clinical decisions in the last two decades. To build robust CAD schemes, we need to develop state-of-the-art image processing and machine learning (ML) algorithms to optimize each step in the CAD pipeline, including detection and segmentation of the region of interest, optimal feature generation, followed by integration to ML classifiers. In my dissertation, I conducted multiple studies investigating the feasibility of developing several novel CAD schemes in the field of medicine concerning different purposes.

The first study aims to investigate how to optimally develop a CAD scheme of contrast-enhanced digital mammography (CEDM) images to classify breast masses. CEDM includes both low energy (LE) and dual-energy subtracted (DES) images. A CAD scheme was applied to segment mass regions depicting LE and DES images separately. Optimal segmentation results generated from DES images were also mapped to LE images or vice versa. After computing image features, multilayer perceptron-based ML classifiers integrated with a correlation-based feature subset evaluator and leave-one-case-out cross-validation method were built to classify mass regions. The study demonstrated that DES images eliminated the overlapping effect of dense breast tissue, which helps improve mass segmentation accuracy. By mapping mass regions segmented from DES images to LE images, CAD yields significantly improved performance.

The second study aims to develop a new quantitative image marker computed from the pre-intervention computed tomography perfusion (CTP) images and evaluate its feasibility to predict clinical outcome among acute ischemic stroke (AIS) patients undergoing endovascular mechanical thrombectomy after diagnosis of large vessel occlusion. A CAD scheme is first developed to pre-

process CTP images of different scanning series for each study case, perform image segmentation, quantify contrast-enhanced blood volumes in bilateral cerebral hemispheres, and compute image features related to asymmetrical cerebral blood flow patterns based on the cumulative cerebral blood flow curves of two hemispheres. Next, image markers based on a single optimal feature and ML models fused with multi-features are developed and tested to classify AIS cases into two classes of good and poor prognosis based on the Modified Rankin Scale. The study results show that ML model trained using multiple features yields significantly higher classification performance than the image marker using the best single feature ($p<0.01$). This study demonstrates the feasibility of developing a new CAD scheme to predict the prognosis of AIS patients in the hyperacute stage, which has the potential to assist clinicians in optimally treating and managing AIS patients.

The third study aims to develop and test a new CAD scheme to predict prognosis in aneurysmal subarachnoid hemorrhage (aSAH) patients using brain CT images. Each patient had two sets of CT images acquired at admission and prior to discharge. CAD scheme was applied to segment intracranial brain regions into four subregions, namely, cerebrospinal fluid (CSF), white matter (WM), gray matter (GM), and extraparenchymal blood (EPB), respectively. CAD then computed nine image features related to 5 volumes of the segmented sulci, EPB, CSF, WM, GM, and four volumetrical ratios to sulci. Subsequently, 16 ML models were built using multiple features computed either from CT images acquired at admission or prior to discharge to predict eight prognosis related parameters. The results show that ML models trained using CT images acquired at admission yielded higher accuracy to predict short-term clinical outcomes, while ML models trained using CT images acquired prior to discharge had higher accuracy in predicting long-term

clinical outcomes. Thus, this study demonstrated the feasibility of predicting the prognosis of aSAH patients using new ML model-generated quantitative image markers.

The fourth study aims to develop and test a new interactive computer-aided detection (ICAD) tool to quantitatively assess hemorrhage volumes. After loading each case, the ICAD tool first segments intracranial brain volume, performs CT labeling of each voxel. Next, contour-guided image-thresholding techniques based on CT Hounsfield Unit are used to estimate and segment hemorrhage-associated voxels (ICH). Next, two experienced neurology residents examine and correct the markings of ICH categorized into either intraparenchymal hemorrhage (IPH) or intraventricular hemorrhage (IVH) to obtain the true markings. Additionally, volumes and maximum two-dimensional diameter of each sub-type of hemorrhage are also computed for understanding ICH prognosis. The performance to segment hemorrhage regions between semi-automated ICAD and the verified neurology residents' true markings is evaluated using dice similarity coefficient (DSC). The data analysis results in the study demonstrate that the new ICAD tool enables to segment and quantify ICH and other hemorrhage volumes with higher DSC.

Finally, the fifth study aims to bridge the gap between traditional radiomics and deep learning systems by comparing and assessing these two technologies in classifying breast lesions. First, one CAD scheme is applied to segment lesions and compute radiomics features. In contrast, another scheme applies a pre-trained residual net architecture (ResNet50) as a transfer learning model to extract automated features. Next, the principal component algorithm processes both initially computed radiomics and automated features to create optimal feature vectors. Then, several support vector machine (SVM) classifiers are built using the optimized radiomics or automated features. This study indicates that (1) CAD built using only deep transfer learning yields higher classification performance than the traditional radiomic-based model, (2) SVM trained using the

fused radiomics and automated features does not yield significantly higher AUC, and (3) radiomics and automated features contain highly correlated information in lesion classification.

In summary, in all these studies, I developed and investigated several key concepts of CAD pipeline, including (i) pre-processing algorithms, (ii) automatic detection and segmentation schemes, (iii) feature extraction and optimization methods, and (iv) ML and data analysis models. All developed CAD models are embedded with interactive and visually aided graphical user interfaces (GUIs) to provide user functionality. These techniques present innovative approaches for building quantitative image markers to build optimal ML models. The study results indicate the underlying CAD scheme's potential application to assist radiologists in clinical settings for their assessments in diagnosing disease and improving their overall performance.

1 Introduction

1.1 Background

According to the trends in the leading causes of deaths in the united states between 1970 to 2002, the age-standardized death rate (per every 100,000 population) from all causes combined decreased by 32% from 1242.2 to 844.6 [1]. The leading causes of death include heart disease, cancer, stroke, accidents, chronic obstructive pulmonary disease, and diabetes. These trends depict a substantial decrease in the rate of deaths among heart disease (-52.152%, from 502.6 to 240.5), stroke (-63.1%, from 151.9 to 56.1), and accidents (-41% from 62.5 to 36.9). Yet, the death rates either increased or relatively remained the same for cancer, chronic disease, and diabetes. The above-mentioned figures may suggest a decrease in death trends per 100,000 population. However, every year the population is increasing, and so is the incidence of many diseases leading to the rise in total deaths. For instance, in the year 2001 in the US, an estimated 1,268,000 new cases of cancer were reported to be diagnosed, among which 553,400 died [2], whereas in the year 2020 in US, an estimated 1,806,590 new cancer incidences with 606,520 deaths [3]. This is still an overall increase in cancer incidence by +42.5% and deaths by +9.5%. This continuous year-over-year growth in the need for medical assistance has fueled the development and adaption of assisted technologies in medicine to expedite and assist radiologists and doctors.

In the past two decades, the rapid growth of technology, the processing power of computers, and the wide availability of cloud infrastructure to digitize medical centers have contributed to major advancements in medical imaging informatics (MII). “MII broadly refers to every aspect of the imaging chain from image creation and acquisition to image distribution and management, image storage and retrieval, image processing, analysis and understanding, image visualization and data navigation, image interpretation, reporting, and communications. The field serves as the

integrative catalyst for these processes and forms a bridge with imaging and other medical disciplines [4]". In current clinical standards, the wide adoption of variants of MII can be outlined into these flows of steps: (i) Patients admitted to the clinical center, (ii) Suggested image acquisition based on the patients underlying condition, (iii) Image preprocessing and segmentation of the region of interest (ROI) from the images, (iv) radiomic feature extraction from the segmented ROI's, background, and global images, (v) building of machine learning models for classification or inference of patient condition, (vi) Integrating Image-based quantitative analysis with clinical and genomic features (markers), (vii) Finally optimal decision making for underlying patient's condition (precision medicine). An example of a radiogenomics system from [4] is outlined in Figure 1-1. In the above example, steps (iii) – (vii), comprising the entire analysis of medical imaging and diagnostic decision making, are termed computer-aided detection and diagnosis (CAD).

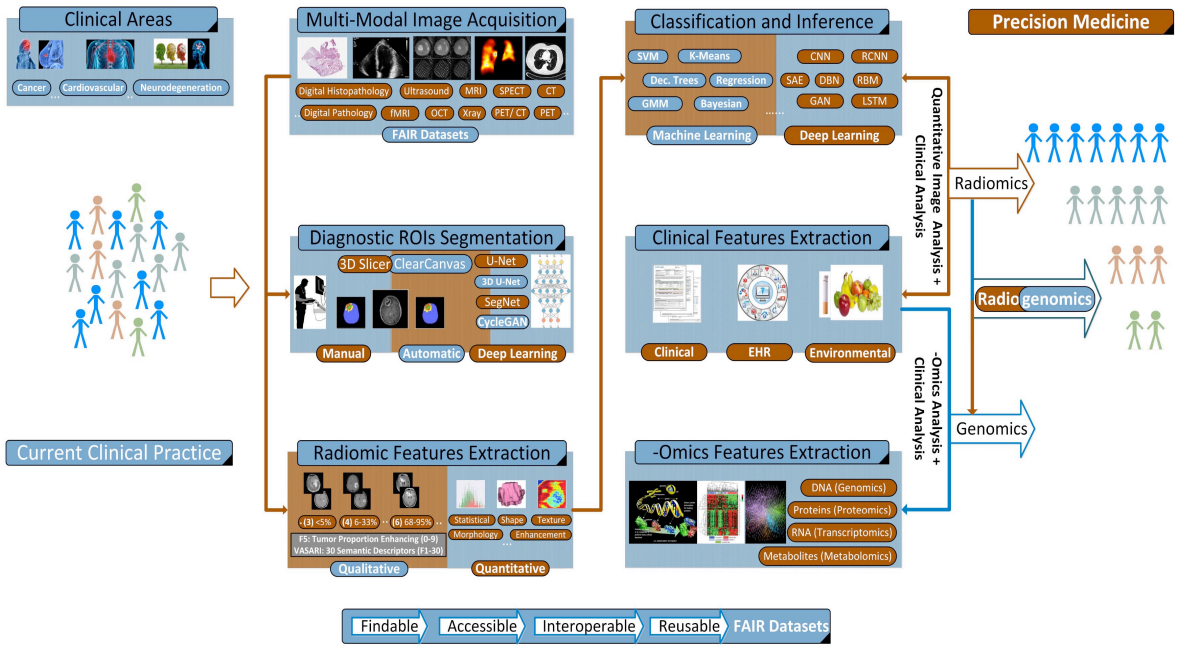


Figure 1-1: Radiogenomics System Diagram: An abstract system diagram demonstrating the use of radiogenomics approaches in the context of precision medicine [4].

The end goal of the CAD schemes is to provide a computer-decision output as a "second opinion" to assist radiologists' in their image readings and diagnostic recommendations. The building blocks and application of CAD schemes can be customized to accommodate multiple imaging modalities and various disease diagnoses. Many CAD schemes were currently adopted in clinical settings and have shown promising results. Some examples include: (i) detection and classification of lung nodules on CT [5], in this study, 16 image readers comprising of both radiologists and residents have all shown to improve their area under the ROC curve (A_z) in detection by using CAD scheme as shown in Figure 1-2. (ii) Classification of pulmonary nodules on low dose CT [6]–[8], a total of 16 radiologists participated in this study, and results were analyzed based on classification of nodules (A_z) for a) radiologists itself without CAD, b) with CAD, and c) CAD itself. The radiologist's performance improved from 0.72 to 0.8 using the CAD, and the CAD scheme independently achieved the best result with an A_z of 0.89, as shown in Figure 1-3. (iii) Detection of Subsolid and Solid Lung Nodules on CT scans using CAD had shown significant improvement on the thick section of CT and a greater significance when thin sections of CT were used [8]. However, some studies show no significant improvement in using CAD for radiologist readings. For instance, in this article [9], a more extensive study was conducted in 90 facilities on 684,956 women with mammographic screenings. The study included 25 (27.8%) facilities adopting CAD for readings. The results show that CAD use during mammography screening is associated with decreased specificity but not with improvement in the detection rate or prognostic characteristics of invasive breast cancer. Additionally, another study [10] reported that CAD systems were widely used in breast cancer screening. However, more of such clinical applications in the other fields are yet to be developed. Thus, there is enormous scope and potential to develop various CAD systems related to multiple health conditions.

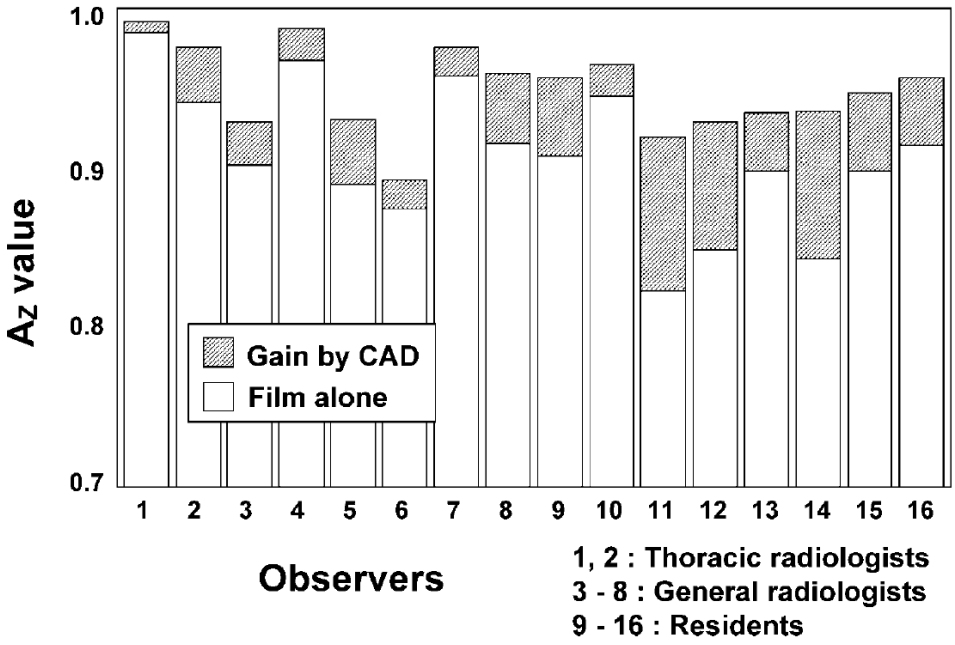


Figure 1-2: Az values without and with CAD for 16 radiologists in the detection of lung nodules on chest radiographs. 60 normal and 60 abnormal with lung nodules of varying subtlety were used [6].

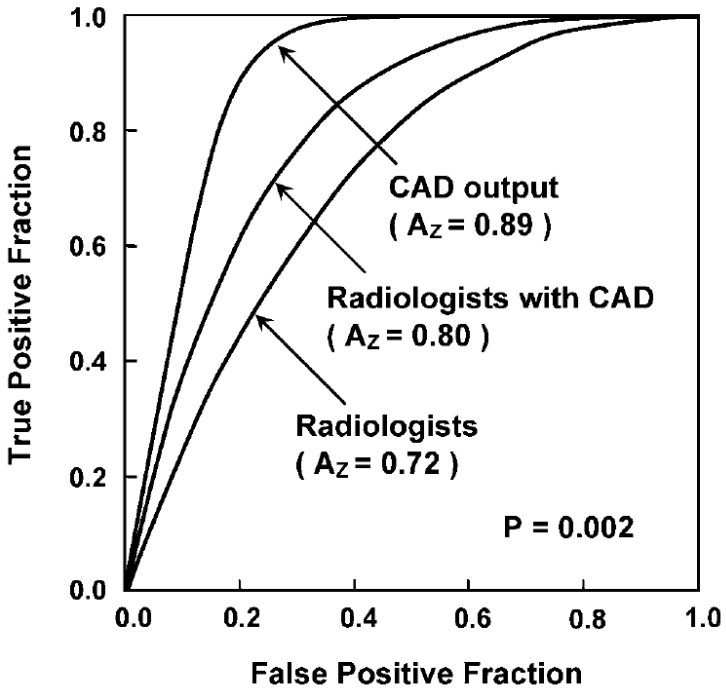


Figure 1-3: Receiver operating characteristic (ROC) curves for distinction between malignant and benign nodules, on chest radiographs without and with the CAD outputs such as those shown in Figure 3. 16 radiologists participated in an observer study in the interpretation of 53 chest radiographs, including 31 primary lung cancers and 22 benign nodules [6].

In the following sections of this chapter, I will first introduce the conventional architecture of a CAD scheme. Then, I review and discuss the recent developments related to each section. Multiple relevant articles will be discussed to highlight the important contributions of the discussed concepts and research steps in these sections toward the overall improvement of CAD performance. I will use research studies that I am a part of as a contributing author or my studies conducted during my Master's thesis. A more detailed explanation of my research during my Ph.D. dissertation period as the first author will be discussed in the later chapters of this dissertation.

1.2 Architecture of Conventional CAD Scheme

A typical computer-aided medical imaging information processing and feature classification system is shown in Figure 1-4. The first step of a CAD begins with image acquisition/ collection from each patient. Then, these images were stored into a secure shareable network system known as a picture archiving and communication system (PACS). PACS systems facilitate image database storing, archiving, and a safe network for accessing image data for end-users like radiologists/ doctors/ and researchers [11]. Next, in the CAD scheme, images were pre-processed to improve image quality, remove artifacts, and eliminate background information to aid in further processing. Then, segmentation schemes were applied to obtain the ROI. Image features from both the segmented ROIs, background, and the global image were then extracted. Optimal image features are identified to reduce the dimensionality of feature space and/or remove redundant features. Finally, statistical prediction models were developed using machine learning systems to make the final decision.

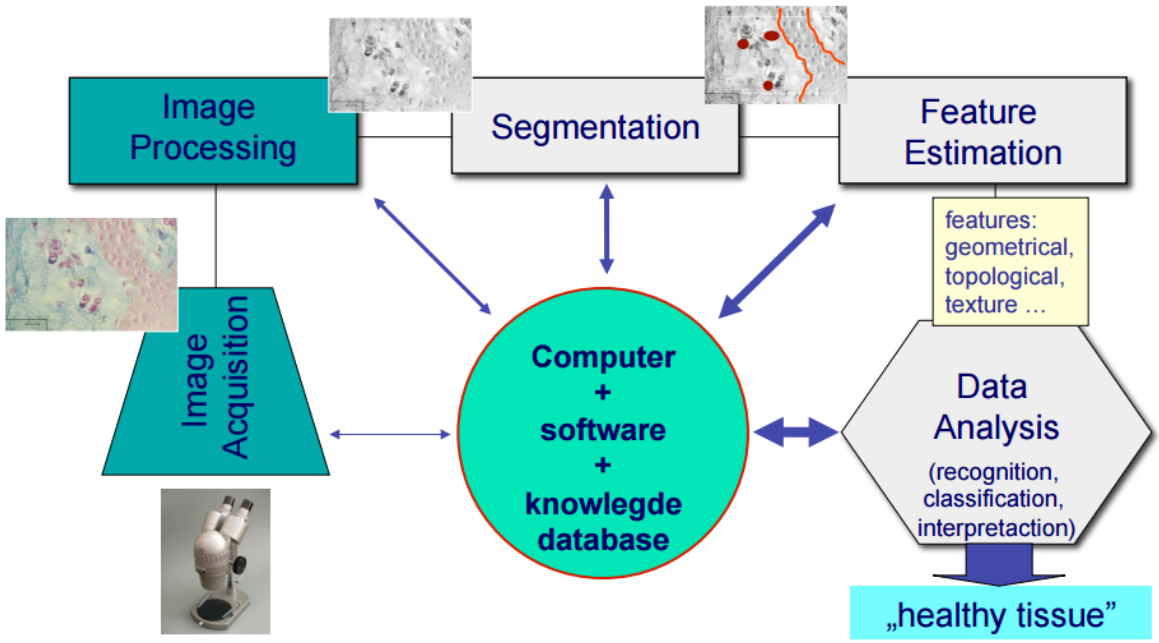


Figure 1-4: Sample architecture of computer-aided medical imaging information processing and feature classification system.

1.2.1 Significance of Imaging Modalities

Medical imaging has become an essential component of the healthcare continuum, from regular screening to early diagnosis, treatment, and follow-up. X-ray is the first two-dimensional medical imaging method introduced in 1895. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) soon followed in the twentieth century. The above three innovations are considered significant milestones in the advancement of medical imaging technology. They all received Nobel Price and are still very widely used in medical practice. Many other new improvements were discovered, including ultrasound, nuclear imaging (single-photon emission computed tomography (SPECT) and positron emission tomography (PET)), interventional molecular imaging, etc. A few critical advancements related to both CT and MRI will be discussed for further discussion.

Many technological advancements were made in CT regarding scan/acquisition speed, slice thickness, decrease in radiation dose, and better image quality. Nowadays, CT scans are widely

available and can be performed in a fraction of a second, covering larger areas. Improvements in image reconstruction techniques and the availability of larger detectors have decreased radiation dosages by more than half, improving image quality. New applications such as CT perfusion (CTP) have revolutionized stroke therapy to detect and quantify stroke regions. In perfusion-based studies, a contrast agent is injected into the cerebral infarct of the patient with stroke onset. If sufficient penumbra is revealed, neuro-interventional treatment can be attempted. In one of my studies, we have used pre-interventional CTP imaging for analysis to identify the cohort of patients who can benefit from EMT in acute ischemic stroke patients. A sample case with large vessel occlusion in arteries leading to stroke is depicted in Figure 1-5 [12]. We can see that CTP used along the angiograph has helped identify the clot and remove the occlusion to resume the blood flow.

Many recent developments were made in MRI too, and previously, one image acquisition was necessary for each type of functional MRI output. But now, multi-contrast MRI imaging from only a single acquisition is made possible, and U.S. Food and Drug Administration (FDA) approved GE Healthcare's MAGiC (MAGnetic resonance image Compilation) software. After scanning, this system can modify image contrast to yield multiple image contrasts, including T1, T2, FLAIR, dual IR, phase-sensitive IR, and proton density-weighted images. This drastically reduces total MRI scanning time and costs and provides clinicians with multiple images to enhance their diagnosis and reduce orders for rescans. Additionally, various contrast-enhanced imaging techniques were also used in breast imaging to identify lesions hidden in the soft tissues and improve overall classification performance. Dynamic Contrast-Enhanced (DCE) breast MRI has improved the cancer detection sensitivities from 40% to 81% compared to traditional mammograms [13]. Hence, breast DCE-MRI has been recommended by the American Cancer

Society as an adjunct screening tool for mammography for women with a lifetime breast cancer risk greater than 20-25% [14]. In one of the studies in our lab on breast DCE-MRI images, we investigated the efficacy of tumor response to chemotherapy using quantitative global MRI image features and yielded an AUC of 0.83 to classify between two classes [15].

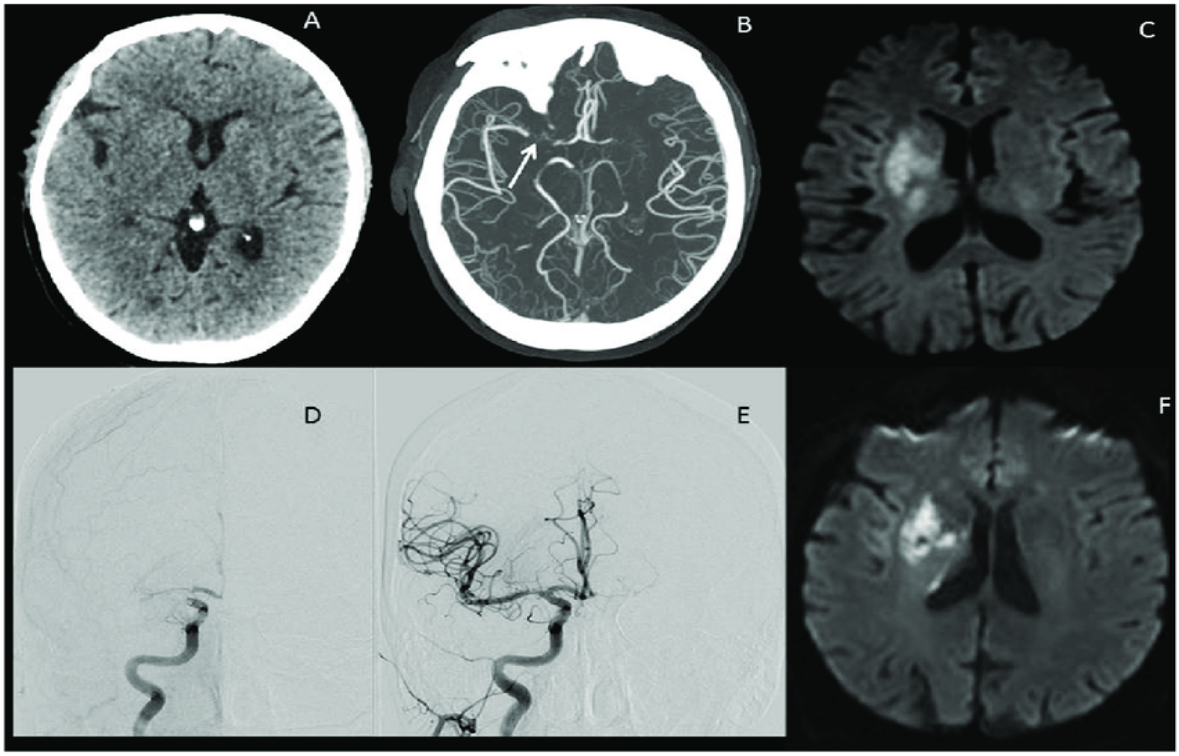


Figure 1-5: Illustrative images of a large vessel occlusion (LVO) stroke patient. Patient was a 67-year-old male presenting 4 h after onset with a full right middle cerebral artery (MCA) syndrome due to right MCA occlusion, NIHSS 14. (A) Emergent head computerized tomography without hemorrhage as a cause of stroke syndrome. (B) Axial maximal intensity projections from CTA showing right MCA occlusion (white arrow). (C) Emergent MRI DWI showing a small established core infarct. On the basis of this combined imaging and clinical data, it was determined that the patient had a large penumbra and small region of established injury and was therefore a good candidate for reperfusion therapy. (D) Anteroposterior view, catheter angiogram. The right internal carotid artery (ICA) injection reveals thrombus at the carotid terminus with only minimal anterior cerebral artery (ACA) opacification seen. Findings are consistent with an ICA-T occlusion. (e) Complete recanalization following mechanical thrombectomy, with full reperfusion (not shown) of the threatened penumbra. (F) 24 h MRI DWI showing arrest of infarct growth following reperfusion of the penumbra. The patient improved to NIHSS 4 by discharge on day 3 post-op. His stroke was determined to be cardioembolic following detection of atrial fibrillation after complete evaluation for cause, and he was free of deficits at 90-day follow-up [12].

1.2.2 Image Preprocessing

Preprocessing is a crucial step in the CAD scheme. There are two main reasons to apply this: to enhance the quality of images to aid further steps (automated segmentation, feature extraction, etc.) and remove noise or unnecessary artifacts from the images. Various filtering techniques from

computer vision are still applicable to grey-scale medical images. This filtering of images can either be performed in the time-domain or frequency domain. A few example methods were as discussed below:

Gabor Filter: This is a linear filter used for texture analysis (edge detection). It is mainly used to analyze whether there is any specific frequency content in the image in a particular orientation in a localized region around the region of interest. This is useful for identifying edges in the images and automatically recognizing text annotation in DICOM medical images.

Mean or Average Filter: A linear kernel-based filter. It is used to replace each pixel value in an image with the average value of its neighbors. It is used to remove pixel values that are unrepresentative of their surroundings.

Adaptive Median Filter: It is a local kernel neighbor-based filter. It is used to remove salt and pepper noise and impulsive noise without causing any distortion to the original images. It is helpful to remove distorted impulsive noise from the original signal preserving the image.

Morphological Filter: Morphological Techniques use a "structuring element" reference template. The elements are used to compare each pixel with its neighbors in the structuring element of the original image to perform filtering. In combination with various structuring elements, these filters can be used to remove/ preserve peaks and valleys from the original images. Many variants of these filters exist, including dilation, erosion, open, close, etc., of the image.

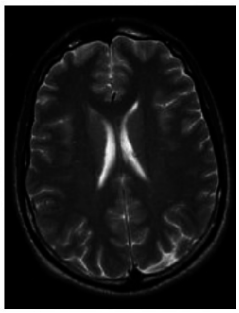
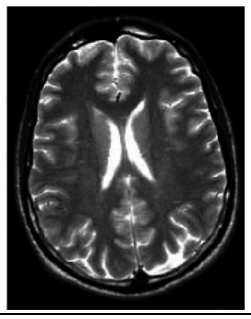
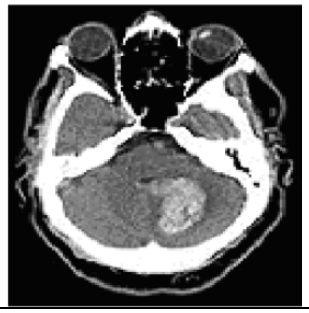
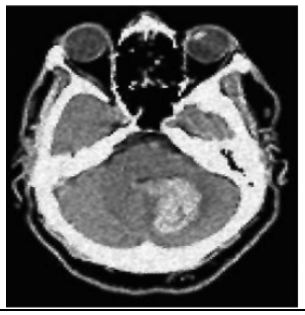
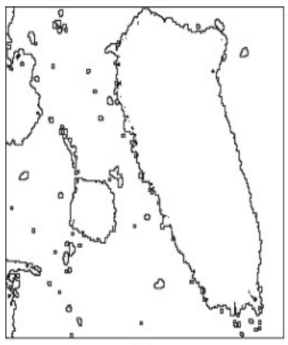
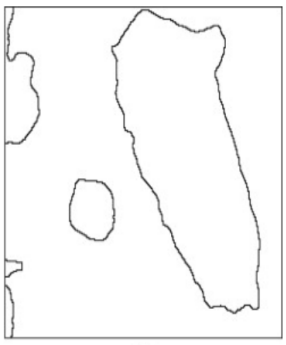
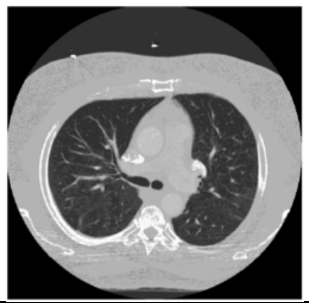
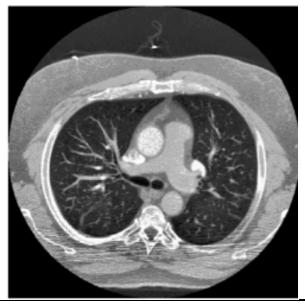
Image Normalization: It is a process of altering the range of pixel intensity values in an image. It is also called Contrast or Histogram Stretching". The main objective of this application is to

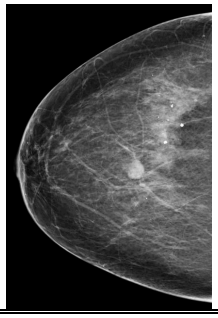
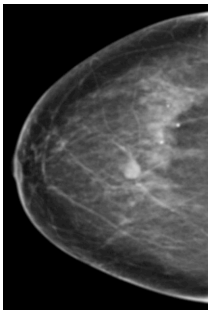
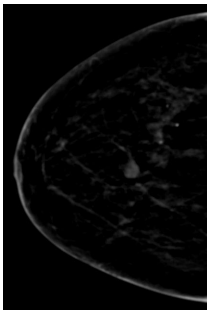
bring all parts of the image or other types of signals into a range familiar to the senses for normal interpretation.

Histogram Equalization: It is used to increase the global contrast of the image, especially when a close contrast value represents the usable data of the image. This equalization helps the intensities in the image to be better distributed on the histogram allowing images to achieve higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values.

Low/High/Band Filters: These filters are frequency domain-based. If the frequency of the image signal satisfies the filter's condition, only then is the signal allowed to pass through the filter. For example, a low pass filter will enable signals less than a selected cutoff frequency and attenuates higher frequency signals in an image. Typically, for the application of frequency-domain filters, the time-domain image is first converted into a frequency-domain (ex: Fourier transform). Then, the filter function is applied (ex: low pass filter), followed by the inverse transform of frequency-domain (inverse Fourier transform), resulting in a filtered time-domain image. In summary, Table 1-1demonstrates several examples of applying different filters as discussed above to medical images.

Table 1-1: Sample images of various filtering techniques with both original and filtered images.

Filtering Technique	Original Image	Filtered Image
Gabor Filter		
Adaptive Median Filter		
Morphological Filter		
Histogram Equalization		

Gaussian Lowpass Filter			

1.2.3 Segmentation of ROI

In computer vision, segmentation can be divided into supervised and un-supervised approaches. In supervised segmentation methods, a prior existing atlas/ knowledge about the ROI and its background is used to perform the segmentation. Whereas in unsupervised methods, segmentation depends entirely on defined segmentation criteria like initial growing condition, growth rate, stop/limiting conditions, etc. Most common supervised segmentation methods include region-growing, adaptive thresholding, contour identification, clustering, etc. Next, I will discuss two articles related to the segmentation of suspicious regions using a unique technique. The first study uses a novel multi-layer topography-based growth segmentation to detect and segment suspicious breast lesions. The second study uses state-of-the-art deep learning architecture for automatic detection and segmentation of four types of retinal lesions in patients suffering from diabetic retinopathy.

1.2.3.1 First Case Study: Concentric Morphology Model for Detection of Masses in Mammography

Multi-level topographic (MLT) is a popular technique used for automated detection and segmentation of suspicious masses. We also used this technique in one of our studies, and it will be explained in more detail in chapter 2. These methods usually follow a defined algorithm for identifying growth and stopping criteria. We will consider a study [16] in which these multi-level

growths of concentric layers were adapted to identify and segment the suspicious breast masses. This study is divided into three stages: (i) breast region segmentation and granulation, (ii) Detection of suspicious focal areas using the MLT technique, and (iii) false-positive reduction.

i. Breast Region Segmentation and Granulation:

Due to large size of mammographs, it is first reduced by a factor of 5 (from 50 to 250 μ per pixel). Then Based on the prior knowledge, the intensity profile of the image is first examined at 11 horizontal sections equally spaced along with the image. By visiting each x-section using fixed intervals of 50 pixels starting from the background and approaching the breast, the breast segmentation module allocates 11 points of high pixel intensity transition close to the breast skin line. Spline interpolation between these advanced points is then used to calculate an initial breast silhouette. After cleaning the background behind this silhouette, histogram analysis is performed on the intermediate image to automatically determine the optimal threshold (Otsu method) for the final segmentation of the breast region. Pixel granulation implements a simple transformation of many pixel intensity levels to a smaller, more manageable number called granule levels. Furthermore, the granulation process ensures that groups of strongly connected pixels in terms of spatial location and intensity range are assigned to the same granule level. A higher granule level denotes strongly connected pixels with brighter intensity values.

ii. Detection of Suspicious Focal Areas Using MLT Technique:

First, the localization of focal areas with suspicious morphology is conducted using granulated breast images. First, all pixels with granule levels equal to or higher than the studied granule level l are identified and analyzed as a separate image. All isolated regions are identified in “focal regions” in this separate image. Although the above method can be repeated further for every

granule level in the breast region, the exhaustive search is computationally costly. Therefore, only the top 50% brightest granule levels are visited at equally spaced granule intervals to reduce complexity. Specifically, the visitation scheme operates as follows. Let us assume that N granule levels are present in the granulated image, where N denotes the highest (brightest) granule level. Starting from the highest granule level 'N' and continuing down to the 0.5*N granule level, the levels are visited with a fixed step of 0.05*N. Consequently, 11 granule levels are visited in total. All seeds detected in an image during the stepwise visitation scheme are collected. However, redundancies in seeds were removed at a minimum distance of 55 pixels threshold. Figure 1-6, Figure 1-7 represent the MCL model and MCL criterion, respectively.

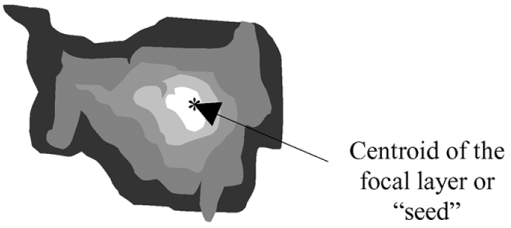


Figure 1-6: The multiple concentric layers mass model.

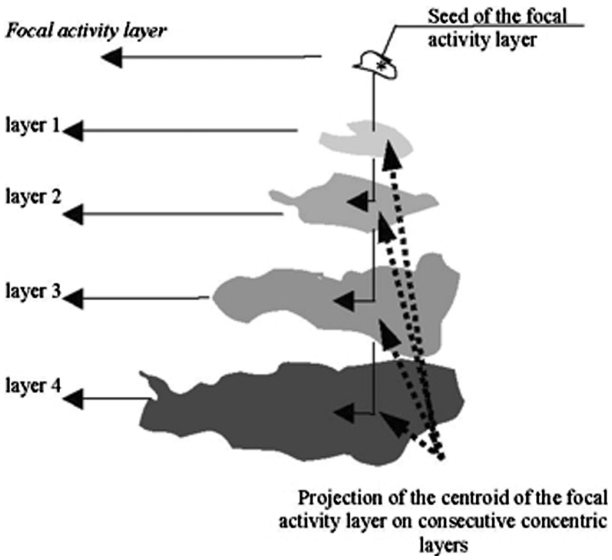


Figure 1-7: Visual representation of the MCL criterion for a focal activity layer detected at granule level l. Four consecutive layers concentric with the focal activity layer is identified at progressively lower granule levels (l - 1); (l -2); (l - 3); and (l – 4).

Next, for each seed, we examine the presence of concentric layers around the seed of progressively lower granule level, as shown in Figure 1-8. Seed regions with higher evolving concentric regions are considered suspicious deviations. In contrast, seed regions with fewer than three evolving concentric layers are eliminated from the list of candidate masses. The MCL is

employed by projecting focal seeds on the consecutive layer while examining the distance between the centroids of every consecutive layer and the projected focal seed on this layer. The layers are considered concentric if this distance is less than 17 mm (68 pixels). Otherwise, it is deemed eccentric, and the search is stopped. If the number of concentric layers found around each focal seed is three or more, the focal area is considered a suspicious mass. The search stops when either one of three conditions is encountered: 1) an eccentric layer is found or 2) an expected granule level is absent, or 3) at least ten consecutive concentric layers are found. A focal region with at least ten concentric layers is considered highly suspicious, and no further search of additional concentric layers is necessary. An example reference case with ROI highlighted along with radiologist markings and MCL layers generated were shown in Figure 1-8.

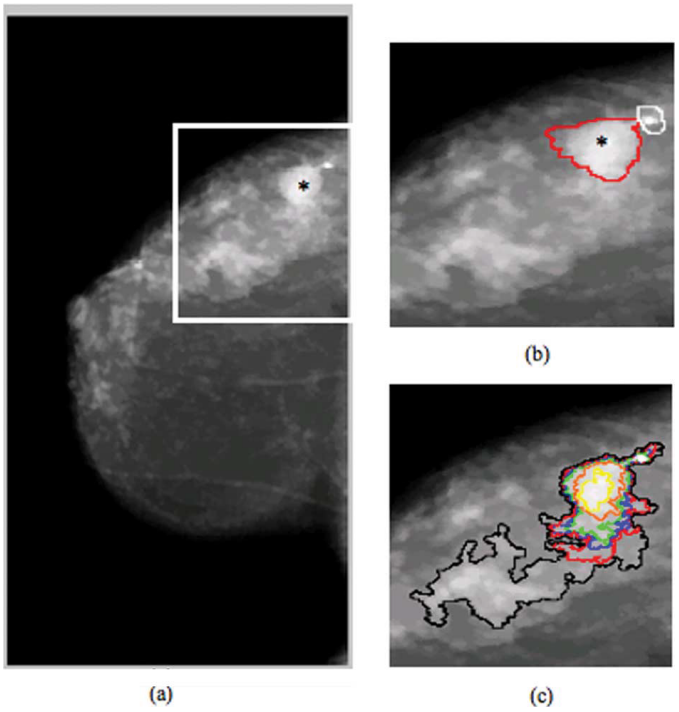


Figure 1-8: (a) Fig with a seed (shown with an asterisk) detected, (b) zooming into the suspicious area to show the physicians' annotations of two true abnormalities present: one malignant mass (outlined in red) and one calcification cluster (outlined in white), and (c) isolated concentric contours surrounding the detected mass seed.

iii. False Positive Reduction

The final MCL stage applies two different criteria for the elimination of false-positive seeds. The first method uses relative incidence with respect to the whole breast region. The second method uses the minimum distance criterion. If two seeds are closer than 75 pixels, then only the seed detected at a higher granule level, including more concentric layers, is kept in the final cue.

Finally, the detection performance of the scheme was evaluated using free-response operating characteristic (FROC) analysis. A CAD detection area was considered as a true positive (TP) if the centroid of its focal activity layer was included in the DDSM annotated area. If the centroid was outside the annotated area, the CAD detection area was considered false positive (FP) even if the extended detected area (i.e., its evolving concentric layers) overlapped with the physician's mass annotation. The results show that malignant masses were detected with 92%, 88%, and 81% sensitivity at 5.4, 2.4, and 0.6 false-positive marks per image. While benign masses showed a false positive rate of 5.0, 1.7, and 0.2 marks per image at the previously reported operating points. The proposed MCL-based suspicious breast masses detection scheme has proven to be very promising CAD for screening mammograms to automatically identify malignant masses.

1.2.3.2 Second Case Study: Applying Deep Learning for Automatic Detection and Segmentation of Retinal Lesions

I am considering another study to discuss automated detection and segmentation using deep learning techniques, where I participated with fellow teammates in the IEEE competition. This study used U-Net's modified deep learning architecture as an inspiration. We performed automatic segmentation of various retinal lesions in Diabetic Retinopathy (DR) patients. Early detection of DR from fundus images is clinically essential for the treatment of DR. The types of retinal lesions analyzed include microaneurysms (MA), hemorrhages (HE), hard exudates (EX) and soft exudates

(SE). We adopt a two-stage training strategy, which randomly extracts negative patches as the training set in the first stage and then selects false-positive patches generated in the first stage for the second-stage training. This paper was submitted to ISBI 2018 diabetic retinopathy segmentation and grading challenge - sub-challenge and was awarded fourth position [17]. This study is divided into three stages: (i) background on DR, (ii) proposed architecture of U-Net, and (iii) fundus image dataset and implementation.

i. Background on DR

Diabetes usually occurs when insulin-producing beta cells in the pancreas are destroyed (type-1) or when various body parts fail to effectively use the insulin produced (type-2). It is a global health issue and one of four priority non-communicable (chronic) diseases (NCD) targeted for action by world leaders of the united nations [18]. Globally, it is estimated that nearly 381.8 million people had diabetes in 2013, and it is projected to increase to 591.9 million (55% growth) by 2035 [19]. This increasing trend is more significant in developing regions (like Asia and Africa) as compared to developed regions [20] (like Europe and North America). Diabetes is chronic in nature, leading to many long-term health complications. Individuals with diabetes have a significant probability of eye damage known as diabetic retinopathy (DR). Since DR is progressive, if untreated in the early stages, it will damage blood vessels of the retina and can potentially lead to permanent blindness [18]. The prevalence of DR among individuals with diabetes is 43%, 28%, 18% in China, the United States, and India, respectively [21], [22]. Thus, appropriate screening guidelines for diabetes patients were needed for early detection and treatment of DR.

A manual interpretation of retinal fundal images by ophthalmologists is widely used to detect DR. During the screening, ophthalmologists observe for retinal lesions of abnormal blood vessels. In general, the early signs of DR are tiny swollen capillaries known as microaneurysms (MA). These lesions further discharge blood into the retina leading to complications such as hemorrhages (HE), hard exudates (EX), soft exudates (SE; cotton wool spots) etc. An illustration of different retinal lesions in DR is shown in Figure 1-9. Furthermore, if untreated will cause an accumulation of fluid in the macula leading to diabetic macular edema (DME). Manual identification of these different types of lesions is crucial for disease grading but is tedious due to the volume of mass screenings. Thus, developing an automated screening tool for the detection and grading of retinal lesions will have potential benefits: cost-effective reproducibility, accessibility to remote places, and effective mass screening schemes.

Many researchers have been using traditional machine learning algorithms with domain-specific handcrafted features for the classification of DR. However, the availability of high computing GPUs lets many researchers explore various deep learning architectures in various fields of computer vision, including medical imaging. In recent studies, early detection of MA using deep neural networks (DNN's) [23] helped in the screening of DR. Additionally, a larger deep learning study on 128 thousand fundus images with annotations of 54 United States licensed ophthalmologists were used for classification into normal and preferable DR [22]. In another study, a novel two-stage deep convolutional neural networks (DCNN's) was built for both lesion detection (stage:1, local network) and automatic grading (stage:2, global network) of DR severity [24]. Thus, in this study, we developed a DCNN using U-Net with adjusted parameters for estimating its feasibility in the automatic detection of various types of lesions (MA, HE, EX, and SE).

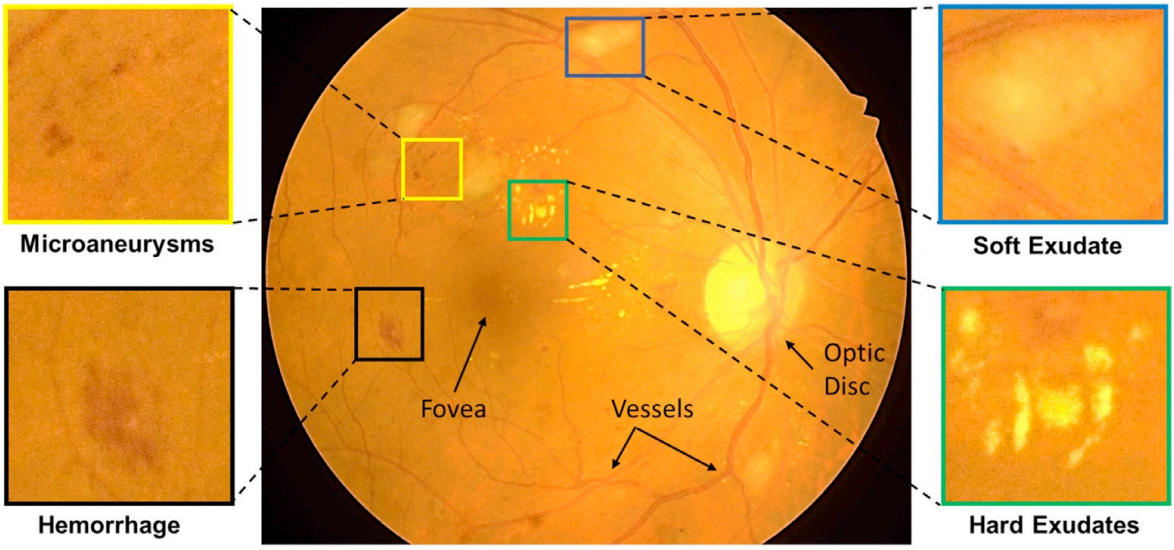


Figure 1-9: Illustration of retinal image (incenter) by highlighting normal structures (blood vessels, optic disc and fovea center) and abnormalities associated with DR: Enlarged regions (in left) MAs, and HEs and (in right) SEs, and EXs.

ii. Proposed Architecture of U-Net

Due to the large size of fundus images and the limitation of GPU memory for training a DCNN, we randomly cropped sub-regions from the fundus images as the input of the networks. We adopt a Fully Convolutional Network (FCN) architecture named U-Net [25], which was proposed for various biomedical image segmentation tasks. Figure 1-10 shows the architecture of the U-Net developed for DR lesion segmentation in our study. The network takes a 380×380 fundus image patch as input and predicts the binary mask of the DR lesion within the 196×196 central region of the input patch. The network consists of a contracting path on the left and an expansive path on the right. In the contracting path, convolutional layers with small kernels (i.e., 3×3) and max-pooling layers are stacked to gradually increase the receipt filed and reduce the spatial resolution of convolved features. In the expansive path, de-convolution operations are applied to increase the resolution of feature maps. Features generated by de-convolution layers are concatenated with the corresponding features with the same resolution in the contracting path to combine coarse features

(from de-convolutional layers) and fine features (from the contracting path). Convolutional layers are then applied to extract high-level representations from the concatenation. The network's last layer is a 1×1 convolution operation with a sigmoid activation function to generate a pixel-level lesion probability map. We adopt the Batch Normalization (BN) technique to improve training efficiency in contracting and expansive paths. All convolution operations adopt a 'valid' padding method to reduce the effects of random cropping.

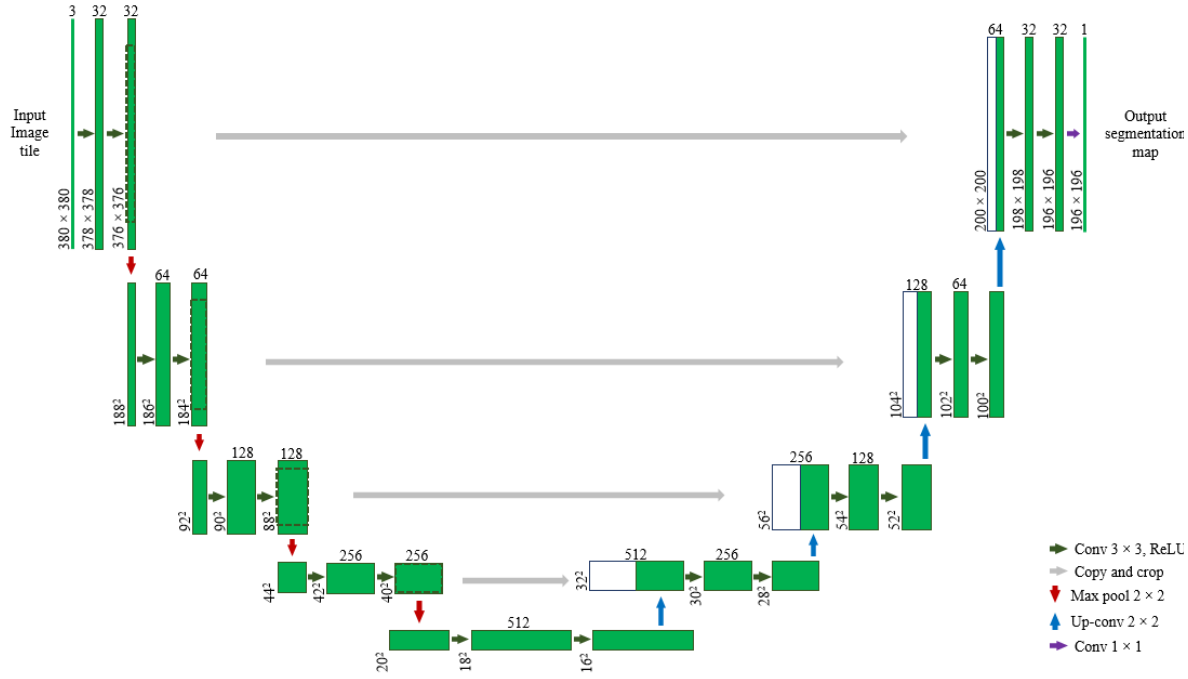


Figure 1-10: Figure 1: Proposed U-Net architecture for patch based retinal lesion segmentation.

iii. Fundus Image Dataset and Implementation

This study's retinal fundus images were retrospectively collected from the existing database at an eye clinic in Nanded, Maharashtra, India. The dataset includes 516 fundal images. All images were acquired using a Kowa VX-10 alpha digital fundus camera with 4288 × 2848 pixels, 50˚ field of view (FOV), and centered near the macula. Measures were taken to ensure quality data

representing each disease stratification of DR and DME is included while collecting the database. The database assembled for lesion segmentation consists of 215 cases, 81 cases have signs of DR, and the remaining 134 were normal cases. Experts and precise annotations examined all the images with signs of DR that were marked for various retinal lesions (MA, EX, HE, and SE). 2/3 of the cases are released with ground truth for training, and the remaining 1/3 are used for testing.

We implemented the U-Net architecture based on the TensorFlow library [26] with an Nvidia GeForce GTX 1080Ti GPU. 380×380 patches were extracted from the original fundus images for training the network. We pre-processed the image patches by subtracting the local mean of each color channel. For data augmentation, each patch was flipped three times (i.e., horizontally, vertically, and both). We adopt a two-stage training process for each sub-type of lesions (i.e., MA, HE, EX, and SE). We extracted positive image patches in the training set for the first stage according to the given ground truth mask. We randomly extracted negative image patches from fundus images with apparent retinopathy and without apparent retinopathy. We trained the U-Net using the extracted patches for 20,000 iterations for each lesion sub-types. The objective function is a standard cross-entropy loss function, and the Adam algorithm is employed to optimize the parameters. Subsequently, we applied the optimized U-Net on the fundus images in the training set and extracted false-positive patches generated by U-Net. We further trained the U-Net for 10,000 iterations using the positive image patches and the false-positive patches as a second stage. In the testing phase, we extracted overlapped image patches using a sliding window and fed the patches into the network to get the corresponding probability maps.

Table 1-2: Performance of U-Net for retinal lesion segmentation.

Lesion subtype	T = 0.1		T = 0.25		T = 0.5		T = 0.75	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
MA	0.640	0.375	0.555	0.504	0.014	0.634	N/A	N/A
EX	0.912	0.525	0.831	0.684	0.699	0.838	0.534	0.944
HE	0.710	0.260	0.546	0.433	0.371	0.635	0.206	0.873
SE	0.455	0.262	0.340	0.411	N/A	N/A	N/A	N/A

While leave-one-out cross-validation is required for the challenge, it is time-consuming for training the network. Therefore, we only split the released training set into one training set and one evaluation set. The training set contains 44 fundus images with apparent retinopathy and 75 images without apparent retinopathy, while the evaluation set contains ten fundus images with apparent retinopathy and 14 images without apparent retinopathy. We optimized the parameters using the training set and evaluated the network on the evaluation and testing sets. So, it should be noted that the results we submitted are not from leave-one-out. Part of the results is obtained by applying the model to the training samples since training the network is time-consuming.

We calculated the pixel-level sensitivity and specificity with different thresholds over the 24 fundus images in the evaluation set. Table 1-2 summarizes the performance of the proposed U-Net for different sub-types of lesions. There are some 'N/A's in the table, which means that the maximum predicted probability is below the threshold.

1.2.4 Feature Extraction and Optimization

Extracting meaningful information from image data relevant to clinicians for deducing underlying phenomena can be significant for diagnosis. Thus, in CADs, feature extraction is performed to capture the radiomic information. These features can be broadly categorized into

three types: geometric-based, Intensity-based, and texture-based. Geometric features are related to the shape of the ROI, and some features commonly used include area, perimeter, compactness, circularity, eccentricity, maximum radius, etc. Next, the intensity-based features are related to the pixel distribution of each ROI, surrounding background, or the global image itself. Some commonly used intensity features include statistical parameters like mean, deviation, skewness, kurtosis, moments, contrast, etc. Finally, texture features capture the underlying textural orientation. Some commonly used types include run-length features, cooccurrence features, local binary patterns (LBP), histogram of oriented gradients (HOG) features, vector quantization generating texture descriptors, and wavelets. Theoretically, many features can be constructed from an image, but the primary purpose of this feature extraction is to reduce the total amount of data represented in an image to a smaller feature profile.

Additionally, having a larger feature dimension can contain much redundant information and may lead to overfitting for decision-making. Thus, optimal feature selection or feature reduction techniques must be employed to reduce the final number of image features. This dimensionality reduction step can also reduce noise and produce more robust learning models.

Feature dimensionality reduction achieved from selecting a subset of the existing features is known as feature selection. The three main approaches to feature selection are embedded, filter, and wrapper approaches. Embedded methods reduce the computational time compared to wrapper methods by incorporating the feature selection in the training process of a classifier. These are very sensitive to the learning algorithm used to set feature subset. Support Vector Machine (SVM) approaches or decision tree algorithms are some examples of embedded methods [27]. Wrapper methods evaluate the utility of feature subsets using the results of a specified classifier. These methods allow for the detection of the possible interactions between variables. A search procedure

within the possible feature subsets space is done. As the number of subsets grows exponentially with the number of features, a heuristic or a sequential selection algorithm is used for search purposes. The two main disadvantages of these methods are the increasing overfitting risk when the number of observations is insufficient and the significant computation time when the number of variables is large [28]. Filter methods do not depend on any classifier but can be considered pre-processing steps based on specific criteria to evaluate features' relevance. One of the main disadvantages of these approaches is that the researchers ignore the interaction between features and hence may be unable to remove redundant features. The most proposed techniques are univariate, which means that each feature is considered separately, for instance on mutual information [29].

Many of our previous studies have focused on feature generation, selection, and optimization to improve the performance of CAD schemes. The following section reviews and discusses our recent study, which focuses on applying a new feature optimization algorithm on medical images to classify breast lesions.

In this study [30], the primary objective is to investigate the feasibility of applying a random projection algorithm (RPA) to build an optimal feature vector from the initially CAD-generated large feature pool and improve the machine learning model's performance for the classification of benign, malignant breast masses. We retrospectively collected a mammographic image dataset consisting of 1487 cases. Of these, 644 were confirmed as malignant lesions, and 843 had benign lesions. The majority of cases have both craniocaudal (CC) and mediolateral oblique (MLO) views in which the suspicious lesions are detected and marked by the radiologists. In contrast, a small fraction of cases have either CC or MLO images. Overall, 1,197 images depicting malignant lesions and 1,302 images showing benign lesions are collected. The centers of lesions were marked

and used as “ground truth” to evaluate CAD performance. Like most CAD schemes using the ROIs with a fixed size as classification targets, we used a 150-by-150 fixed size ROI centered around lesion markings for generating our initial feature pool.

The features extracted can be categorized into three types: (i) statistical, (ii) geometrical, and (iii) textural (which includes gray-level run-length matrix (GLRLM), gray level difference matrix (GLDM), gray-level co-occurrence matrix (GLCM), wavelet) features. In summary, 181 features were computed either from the global ROI images or the segmented ROIs.

Before using RPA to generate an optimal feature vector from the initial image feature pool, we first normalize each feature to make its value distribution between [0, 1] to reduce case-based dependency and weight all features equally. Thus, for each case, we have a feature vector of size d, which is valuable to determine that case based on the extracted features as a point in a d dimensional space.

For evaluation, we adopted the support vector machine (SVM) to train these new RPA-based features to predict the likelihood of lesions being malignant. Additionally, the leave-one-case-out (LOCO) based cross-validation method was employed to increase the size and diversity of training cases and reduce the potential bias in case partitions. The results were then analyzed using both Area under ROC curve (AUC) and classification accuracy. Finally, the performance of the proposed RPA method was compared with other existing methods like PCA, Chi2, and NMF. Figure 1-11 shows the caparison of ROC curves generated by various models to classify malignant and benign breast cases or lesions. Table 1-3 summarizes multiple parameters, including classification accuracy, sensitivity, specificity, and Odds Ratio of the models.

Table 1-3: Summary of the lesion case-based classification accuracy, sensitivity, specificity, and odd ratio of using 5 SVMs trained using different groups of optimized features.

Feature sub-group	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**Odds Ratio**
Original features	69.3	62.0	75.0	4.85
NMF	72.4	63.1	79.5	6.61
Chi2	70.9	63.0	77.1	5.67
PCA	72.8	68.0	76.6	6.87
RPA	**75.2**	**70.2**	79.0	**8.86**

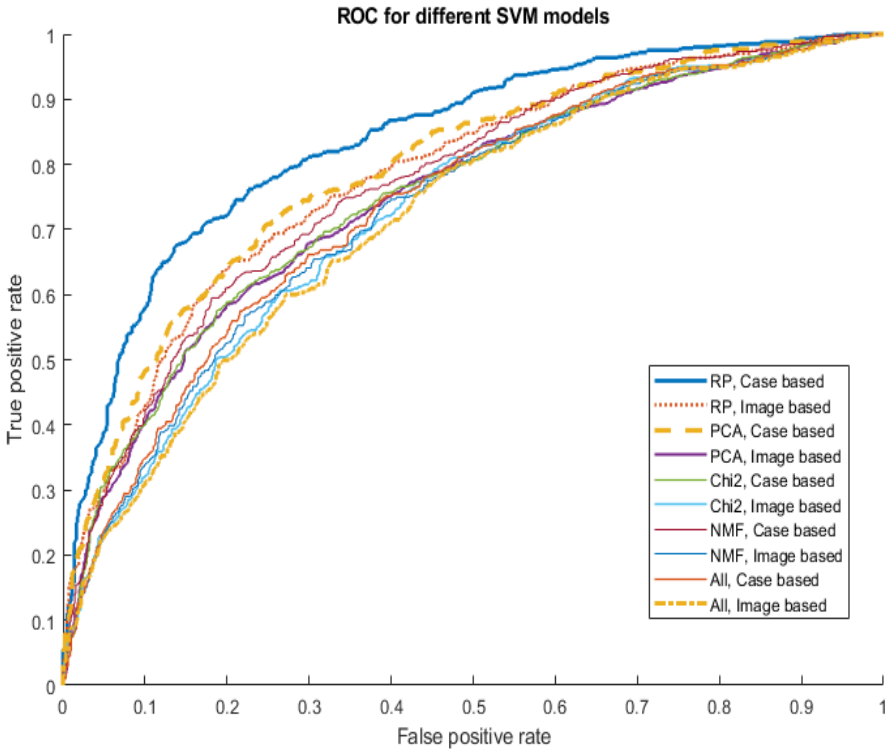


Figure 1-11: Comparison of 10 ROC curves generated using 5 SVM models and 2 scoring (region and case-based) methods to classify between malignant and benign lesion regions or cases.

1.2.5 Content-Based Medical Image Retrieval

In the clinical practice of reading and interpreting medical images, clinicians (i.e., radiologists) often refer to and compare similar cases with verified diagnostic results to detect and diagnose

suspicious lesions or diseases. However, identifying similar reference cases from a large and diverse clinical database or repositories is a quite difficult task. Therefore, developing an automatic and effective medical image retrieval (MIR) system is required to aid clinicians and/or radiologists in browsing the large datasets in clinical PACS. Content-based image retrieval (CBIR) schemes are the most appropriate and reliable approach to automatically retrieve clinically relevant cases and related medical images to assist clinicians in their decision-making process [31]–[33]. The general architecture of CBMIR is shown in Figure 1-12.

CBIR has been one of the most active research areas in computer vision [34]. CBIR refers to the recall of images from the database relevant to the query, using information derived from the images. Many CBIR schemes related to natural images (color) have been proposed and commercially accepted and/or available on the internet [35]. Nevertheless, the application of CBIR models for either biomedical research or routine clinical settings is limited. There are many possible reasons for the delay of such a popular technique (CBIR) to solve medical problems [36] (content-based medical image retrieval (CBMIR)). Some of the reasons are as listed here (but not limited to): (1) lack of collaborations between medical and engineering experts (due to data ownership, privacy, etc.), (2) lack of adequate representation of medical content (mostly grayscale images) by low-level mathematical features, (3) lack of thorough evaluation of CBIR model performance in health care settings, and (4) lack of availability and/or acceptance of appropriate CBIR tools for clinicians to experiment, etc. We will discuss one article of CBMIR in detail from my general examination described with a relevant application to medical image data.

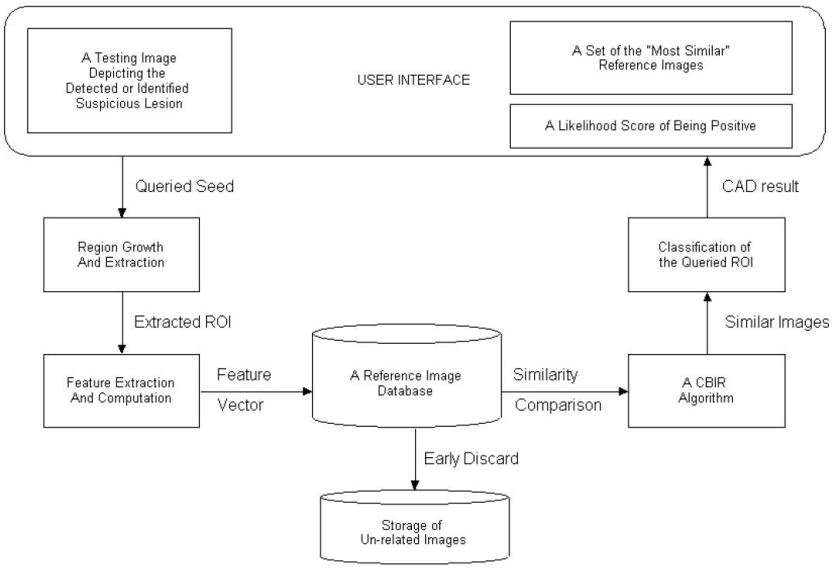


Figure 1-12: General framework of a sample CBMIR of mammograms.

Real-time image retrieval is crucial in using CBMIR systems in clinical diagnosis. For example, in the automatic analysis of massive histopathological images, the images are too big and often represented with over 10,000-dimensional image feature vector. In [37], a CBIR system is designed to retrieve images and their associated annotations from a networked microscopic pathology image database based on four types of image features. Additionally, another study in [38] proposed a CBIR system using the multi-tiered approach to retrieve microscopic images, allowing multi-image query and slide-level image retrieval to protect semantic consistency. However, scalability is the key factor in CBIR for medical image analysis. The above-discussed CBIR systems have only been tested on a relatively small number of cases. Thus, in an attempt to design a computational and scalable retrieval algorithm, [39] proposed a hashing-based approach to analyzing histopathological images. This study used a state-of-the-art kernelized and supervised hashing (KSH) method to achieve optimal performance and moderate training costs. The fundamental idea of KSH is to bridge the gap between the low-level hash code similarity and the

high-level semantic (label) similarity by use of supervised training. The proposed architecture of the KSH-based CBIR system is as shown in Figure 1-13.

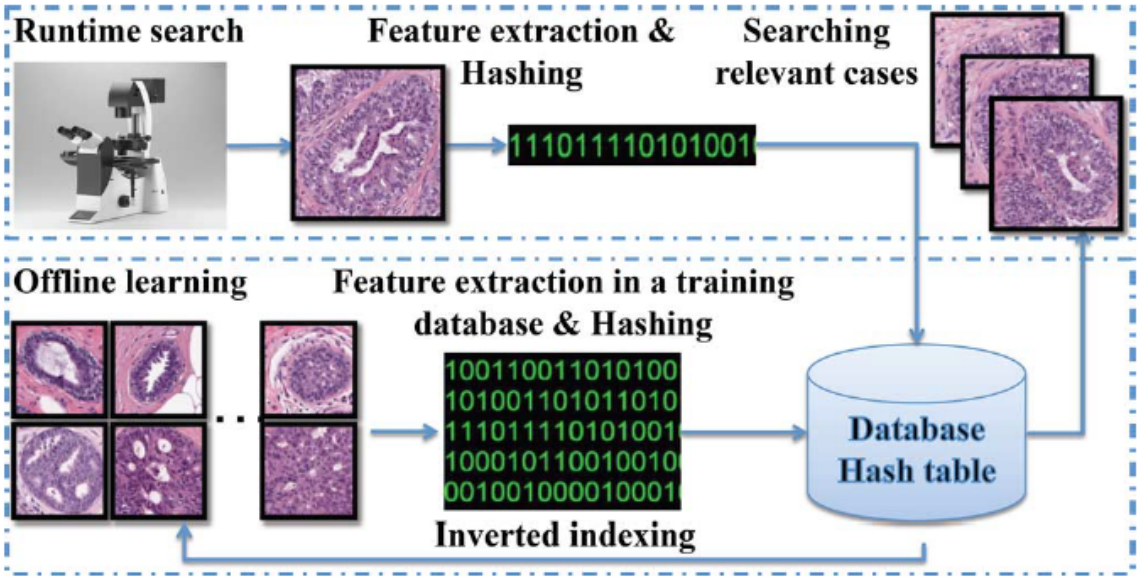


Figure 1-13: Proposed architecture of the kernelized and supervised hashing (KSH) based CBIR system.

During training, high-dimensional visual features are extracted from histopathological images based on scale-invariant feature transform (SIFT). These features represent textual and appearance information and are quantized with a bag of words. It can provide an informative description of cell appearance and robust to subtle staining color changes. Then, a hashing method is used to compress these features into 48 bits of binary codes allowing easy mapping into a hash table for real-time search. A similar step can be used to reduce the high-dimensional feature vector to obtain small binary codes during a run-time query. Finally, using a hash table, searching for nearest neighbors can be achieved in a constant time, irrespective of the number of images in the reference database. Few retrieved sample results generated using the proposed KSH-based image retrieval system of histopathological images of biopsied breast tissues are shown in Figure 1-14.

To compare the performance of this study, the authors also compared this new KSH-based CBIR scheme with other classifiers that have been previously developed and tested for histopathological image analysis. Specifically, the conventional k-nearest neighbor (KNN), principal component analysis (PCA), neighborhood components analysis (NCA), and graph embedding were modeled and compared. The parameters that were tested are (1) precision for the top 10, 20, and 30 results, (2) memory cost of training data, and (3) query time for all test images. A comparison of the performance of the proposed system with other approaches is shown in Table 1-4. Many different modified versions of hashing-based large-scale image retrieval techniques (like sparse hashing, vocabulary tree-based hashing, composite anchor graph hashing) [40]–[42] have also been proposed in recent years and have shown encouraging results.

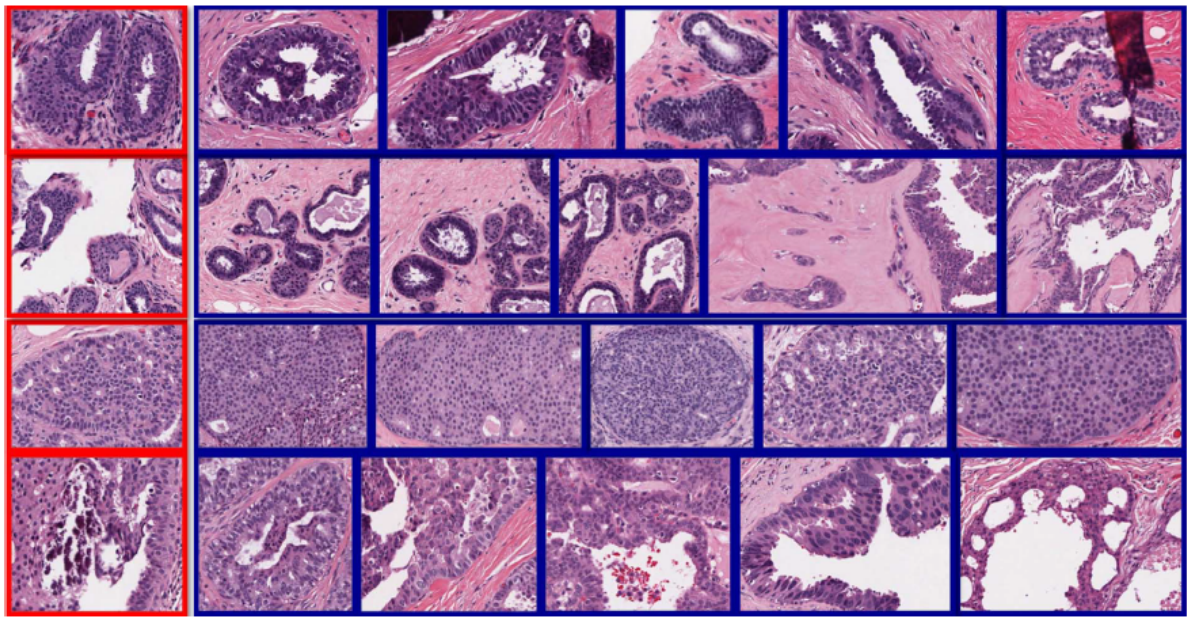


Figure 1-14: Sample results of the KSH based image retrieval system of histopathological breast tissues. First two rows are benign, the last two are actionable.

Table 1-4: Comparison of performance of the proposed KSH based CBIR system with other approaches.

	kNN		PCA		NCA		Graph Embedding		Proposed KSH	
	benign	actionable	benign	actionable	benign	actionable	benign	actionable	benign	actionable
P@10	0.779	0.687	0.762	0.705	0.799	0.697	0.672	0.487	0.836	0.830
P@20	0.773	0.653	0.758	0.681	0.800	0.689	0.673	0.486	0.839	0.829
P@30	0.770	0.631	0.755	0.667	0.800	0.685	0.670	0.480	0.837	0.833
STD	0.024		0.028		0.020		0.012		0.011	
Time	15.77		10.07		10.04		10.03		< 0.01	
Memory	134.58MB		0.65MB		0.65MB		0.65MB		0.01MB	

1.3 Organization of the Dissertation

During the past several years of my study at the University of Oklahoma for my Ph.D. degree, I participated in multiple projects related to the development of various CAD schemes. These studies involved various applications of machine learning concepts in medical imaging informatics. I have collaborated with both researchers and medical professionals, which helps me better understand the current clinical challenges and the need to develop new CAD schemes, emphasizing assisting clinical decision-making. These studies included interactive and visually aided GUI application tools that researchers can use to observe and assess critical steps to help develop robust CAD models.

This dissertation reports five research studies addressing various challenges in developing CAD schemes of medical images, including (i) proposing novel segmentation of ROIs, (ii) extracting and optimizing radiomic features, and (iii) applying several machine learning algorithms to perform classification and segmentation tasks. Chapter 2 will first briefly introduce the current challenges related to each suggested research study and the research hypothesis to

address or solve each challenge. Then, a more detailed assessment of these studies separated into individual sections will be provided in the subsequent Chapters 3 - 7.

Specifically, Chapter 3 reports the investigation of applying CAD technology to a relatively new imaging type (contrast-enhanced digital mammograms in collaboration with Mayo Clinic) to classify suspicious breast masses. By collaborating with neurologists and medical residents at both University of Oklahoma Health Science Center (OUHSC) and the University of Texas Southwestern Medical Center, I developed and tested several new CAD schemes of brain CT images. Chapter 4 reports developing novel quantitative image markers to predict the prognosis of acute ischemic stroke. Chapter 5 reports another CAD scheme developed to predict multiple clinically relevant measures for both short-term and long-term outcomes in patients after aneurysmal subarachnoid hemorrhage. Chapter 6 reports another interactive CAD software tool developed for quantitative estimation of intracerebral hemorrhage. Next, Chapter 7 reports an assessment study to observe the correlation between traditional radiomic-based features and automatically generated deep learning features to classify the suspicious breast lesions. Finally, Chapter 8 summarizes the role of these proposed CAD schemes and discusses their application potential and the prospective future work.

2 Research Objective and Hypothesis

2.1 Current State and Challenges in Developing Robust CAD Schemes

The development of CAD schemes in MII is still emerging and has a great potential to help radiologists in their clinical practice. During the last couple of decades, CAD schemes in commercial and research centers have attracted great interest and adaption. Many commercialized CADs have been widely adopted in many critical fields of medical ailments, including (i) breast cancer screening for detection of suspicious masses, (ii) detection of lung nodules on low dose CT, (iii) in the field of brain stroke or hemorrhagic analysis of quantification or segmentation of edema and blood volumes and many other medical applications. Commercialized CAD systems were widely tested and used clinically as a "second reader" to assist radiologists in interpreting medical images. This trend of CAD development and implementation has been accelerated due to many catalysts like (i) improved imaging technologies, (ii) growth in processing speeds, (iii) evolution of the concept of radiomics which depict phenotype features that are highly associated with genomic and radiologic markers, (iv) advancements and application of deep learning architecture, (v) more research interest in the field of machine learning focusing on medical imaging informatics, etc.

However, few previous studies have shown and well-investigated a gap between CAD and radiologists' performance improvement. The article [9] is an extensive study on CAD for breast cancer screening as a second assisted reader for radiologists with decreased detection rate and specificity performance. In our experience, major challenges for CAD shortcomings are (i) lack of visual assisting tools, (ii) identification of more radiomic related image features to improve the confidence of radiologists, (iii) selection of optimal feature pool for building a more robust machine learning models, (iv) creating more transparent deep learning models to visualize

intermediary steps, etc. In the previous chapter, we have discussed several recent publications focused on key steps in building CAD systems. Specifically, the articles [5], [7], [8] have highlighted the contribution of CADs towards improving radiologists' performance in image readings. The role of new improvements in medical imaging technologies and their contribution to image reading and CAD application were summarized in articles [4], [6], [43], [44]. The purpose of various image preprocessing schemes was introduced, and the implementation of key imaging types was analyzed.

Additionally, in the development of CAD schemes, detection and segmentation of various suspicious regions play a vital role in decision-making. Thus, we discussed two very different articles in this report and developed them for automatic detection and segmentation. In the first study [16], a multi-layer topographic-based concentric morphology model was explained in detail to detect masses in mammograms. The study introduces a novel scheme, which uses granulation to identify multiple concentric growth layers following a certain defined growth criterion, thereby identifying potential suspicious seeds/ masses. In the second study [17], I discussed our article on retinal lesion segmentation in diabetic retinopathy patients. We developed a deep learning architecture inspired by the U-Net to build a model for identifying and segmenting four different retinal lesions. The study examined the feasibility of our system in categorizing different retinal lesions from one another and its pixel-level sensitivity and specificity for their segmentation. Next, the report reviews the importance of focusing on optimizing the performance of the CAD schemes with different feature extraction and data reduction coupled with machine learning schemes. We took one of our studies [30], explaining the application of a random projection algorithm for optimizing the feature reduction step in our study conducted on the classification of benign and malignant breast lesions. The study indicated that the classification performance of the random

projection algorithm is significantly better for feature optimization than other standard techniques like the principal component analysis.

Next, in [41], the hashing-based image retrieval technique was evaluated to retrieve large-scale histopathological breast tissue images. The proposed hashing technique represented a high-dimensional feature vector (10,000) in only 48 binary bits, effectively representing the feature information. The proposed hashing-based CBIR results were also compared with other popular feature reduction-based techniques, which significantly improved parameters such as average precision, time complexity, and memory requirement.

Nevertheless, despite the encouraging results of various recent studies in CAD, many limitations exist and need further exploration. First, many of these CAD studies are conducted on relatively small datasets. Thus, it is not easy to estimate its efficacy in a large population-based environment. The challenge here is to access larger image datasets, which is not practical in many instances due to data privacy protection and institutions' unwillingness to make data public to research. Second, more collaborations need to be conducted between CAD researchers and radiologists to bridge the gap and develop robust CAD schemes considering the visual markers radiologist observe in readings. Third, emerging radiomic features, optimal features selection, and reduction techniques must be further investigated. Finally, any level of automation/ generalization of CAD schemes may not be enough to gain the trust of the radiologist in adapting them as assisting tools. Thus, study-specific customizable visual tools based on radiologists' recommendations need to be incorporated into CAD schemes.

In summary, the above-discussed literature search and review process has helped me better understand different aspects, progress, and challenges of developing CAD schemes in the field of

medicine. It also encouraged and guided me to identify the new research directions and concerns related to improving and developing new CAD schemes. Thus, based on my understanding, to develop a robust CAD model, I first lay out the complete workflow of the current application to identify the major challenges that hinder the efficacy of its implementation. Then, I examine and compare various possible strategies concerning different aspects of CAD architecture to develop an optimal strategy or scheme to solve the research problems. In the following section, I will present some current challenges/objectives related to developing CAD schemes. Then, I will provide my research hypothesis and implement techniques used to solve each research problem.

2.2 Investigating New Imaging Modality in Developing Robust CAD Scheme

2.2.1 Background

Various radiographic imaging modalities are currently being used to screen internal body structures. These modalities are categorized based on the technique used to generate these images. The most commonly used imaging modalities include ultrasound, x-ray, CT, MRI, and positron emission tomography (PET). Ultrasound devices are portable, real-time to capture anatomical and functional images. At the same time, CT is a widely available and fast process that achieves greater clarity using multiple projected x-ray sources and detectors. Additionally, even though time-consuming, MRI uses hydrogen atoms in the body and is very useful for depicting physiology and anatomy. Finally, PET scanners use the targeted injection of radioactive substances before imaging to detect and predict the prognosis of cancer and other diseases.

Many of these techniques offer complimentary information, so in practice, they are often used in combination to observe diverse patterns related to various organs' anatomical, physiological, and functional aspects. For instance, it is very common for patients with brain injury to undergo an immediate CT angiogram at admission. Then followed by contrast injected CT and MRI (T1,

T2, FLAIR) scans. The initial CT allows observing the broad anatomical structures like blood vessels, bone structures, internal bleeding, etc. The later scans offer specific functional features related to patients' response to the treatment. For another example, it is common to use full-field digital mammograms (FFDMs) in a large population in breast cancer screening. However, the people with high risk are later screened using advanced imaging like MRI and dynamic contrast-enhanced (DCE)-MRI.

Even though FFDM and DCE-MRI are the two most common imaging modalities in breast cancer screening, they have their limitations. FFDMs have relatively low sensitivity and specificity due to the fibro-glandular tissue (FGT) overlap in the two-dimensional projection. DCE-MRI is costly, time-consuming, and has lower specificity. Additional challenges in accurate detection and segmentation of suspicious breast masses exist in both modalities due to the overlap of dense FGT affecting the robustness in building CADs. Thus, there is a need to investigate and develop another novel imaging modality that can be fast, accessible, and has reasonable sensitivity and specificity ratings.

2.2.2 Hypothesis and Proposed Approach

A novel alternate imaging modality, namely contrast-enhanced digital mammography (CEDM), is being widely investigated to overcome the aforementioned disadvantages. Using this new imaging modality, I developed a CAD scheme to observe its feasibility in classifying suspicious breast masses. The idea here is that as CEDMs are acquired at two different x-ray energy levels (low energy (LE) and high energy (HE)) after injecting the contrast agent, it allows for observing the morphology and vascular enhancement of suspicious lesions. The contrast agent flows through the blood vessels, and its permeability is more around malignant masses than benign masses. Additionally, when the two images obtained for the CEDM technique are logarithmically

subtracted the difference in the permeability can be further enhanced to eliminate the overlapping effect of FGT tissues and achieve images similar to that of MRI. Thus, this new technique takes the maximum advantages of both FFDM and DCE-MRI while reducing their shortcomings.

In this study, I hypothesized that the two CEDM images offer complimentary information that can be used to improve the performance of classifying suspicious breast lesions. Thus, I proposed to investigate a new and optimal approach to develop a fully automated CAD scheme of CEDM images with two unique characteristics. First, dual-energy subtracted (DES) images enhance breast lesion regions while removing and/or suppressing normal parenchymal tissues that overlap or surround the lesions. Thus, segmentation of breast lesion regions from DES images becomes much more accurate and robust. Second, the lesion density heterogeneity information is predominantly observed in the LE images compared to that of the DES images, allowing the capture of useful radiomic image markers. Thus, the study's objective is to test my hypothesis of using the DES images for segmentation and LE images to generate optimal image markers in classifying suspicious breast lesions.

The study's unique contribution is that we investigated and tested a new approach to developing the first automated CAD scheme of breast lesion classification using CEDM images. Study results demonstrated that LE and DES images generated from CEDM contain complementarily valuable information. This study helped establish a solid foundation for us and/or other researchers to continue developing and optimizing novel CAD schemes of CEDM images with improved performance in future studies. The details of this study will be presented in chapter 3.

2.3 Developing Quantitative Image Markers to Predict Disease Prognosis

2.3.1 Background

The primary objective of developing radiographic image markers is to extract clinically relevant information for the imaging data related to disease detection, diagnosis, and/or prognosis. These image features are carefully handcrafted to mimic what a radiologist or healthcare professional observes while reading the image scans. However, it is very challenging and problem-specific to design these image markers.

In assessing the acute ischemic stroke (AIS) patients with large vessel occlusion (LVO), the current clinical trials use the estimation of cerebral infarct core and salvageable radiological brain tissue "at-risk" for infarction [45]. These clinical image markers to select the patients for endovascular mechanical thrombectomy (EMT) depend on the arterial input function (AIF) and venous output function (VOF) to provide an estimate regarding several critical makers. These markers include cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and time to peak (TTP or Tmax) [46]. However, these techniques cannot capture microcirculatory dynamics of contrast flow through the brain parenchyma. Additionally, these estimations of stroke volumes from the CTP images are often inaccurate compared to those depicted in the MRI-based DWI imaging [47]. This qualitative image assessment of stroke severity using the above-mentioned clinical image markers lacks quantitative assessment and large inter-reader variability [48].

Moreover, recent studies on aneurysmal subarachnoid hemorrhage (aSAH) patients show that early brain injury (EBI) and delayed cerebral ischemia (DCI) that results from several pathophysiological processes are major determinants of mortality and morbidity associated with aSAH patients [49]. However, the absence of objective radiological image biomarkers limits its

role in assessing disease severity and predicting prognosis. The commonly used modified Fisher scale (mFS), although easy to use and able to predict cerebral vasospasm and DCI, it is subjective with high inter-rater variability [50],[51]. Currently, qualitative tools are being used to estimate the blood to predict DCI occurrence and long-term clinical outcomes but suffer operator dependence [52]. Additionally, quantification of cerebral edema using semi-automated or automated segmentation of sulci volumes has shown to be an essential marker for analyzing EBI after aSAH [53],[54].

2.3.2 Hypothesis and Proposed Approach

Thus, to address the above two clinical challenges in the patients suffering from AIS and aSAH, I developed two separate CAD schemes that focus on developing fully quantitative image markers to address the associated clinical problems. In the first study related to AIS patients, different from the existing techniques, I tried to observe the bilateral asymmetry in blood flow patterns to generate image markers. Whereas in the second method, automatic labeling and quantification of several clinically relevant volumes are computed to build image markers. The two different tasks are supposed to evaluate the performance of my proposed schemes. The purpose of the first task is to predict the prognosis of AIS patients. In comparison, the purpose of the second task is to predict various clinical complications that occur after the incidence of aSAH.

For the first task, I investigate the feasibility of developing new quantitative image markers computed from CTP images at an early diagnosis stage to predict AIS prognosis. For this purpose, we developed a new CAD scheme with several novel image processing algorithms to segment the contrast agent enhanced blood volumes in bilateral cerebral hemispheres of the brain, generate cumulative blood flow curves and then compute asymmetrical blood flow related features in two brain hemispheres. Then, image markers based on the best single feature and ML models fused

with multi-features are developed and tested to predict clinical benefit or outcome in a group of AIS patients undergoing EMT for LVO.

This first study has several unique characteristics and contributions. First, we apply several novel image processing algorithms to develop a new CAD scheme that can be applied to real clinical images with varying imaging scanning conditions. Based on our literature search, no similar CAD schemes are available to date. Second, due to the potential presence of unilateral blood clots (LVO) in AIS patients, the transit time and velocity of the blood contrast flow rate (wash-in and wash-out) may vary between two hemispheres of the brain. Thus, asymmetrical blood flow rate or pattern in two brain hemispheres provides a potentially useful image marker to predict AIS prognosis. Third, CAD demonstrates the feasibility of identifying and applying radiographic image features or markers computed from brain CTP images to phenotype AIS patients and predict their prognosis potentially. The study shows promising results when using the proposed CAD scheme and ML model to a set of diverse clinical cases with different Modified Rankin Scale (mRS) distribution and varying CTP imaging scanning protocols. The details of this study are presented in chapter 4.

In the second task, the hypothesis is motivated from two clinical studies [53], [55] to automatically segment and quantify at a pre-determined level above the lateral ventricle to perform a volumetric assessment of sulci, white matter (WM), gray matter (WM), and extra-parenchymal blood (EPB). Applying an automated program to detect and segment brain CT images has the potential to provide a standardized and unbiased estimation of targeted radiological image parameters. Thus, in this exploratory study, the objective is to develop and test a CAD scheme is to accurately quantify image features that pathophysiologically contribute to short-term (in-hospital) and long-term clinical outcomes. Successful development of such radiological image

biomarkers will help in the early prediction of possible clinical outcomes of aSAH patients. The study results indicate the significance of both the admission (first) and day 10-14 (last) CT scans in their predictive capability of assessing both short-term and long-term clinical measures. The details of this study will be presented in chapter 5.

2.4 Designing Novel Segmentation Algorithms to Build Effective CAD Schemes and/or Interactive Software Assisting Tools

2.4.1 Background

Accurate segmentation of ROI is a critical step in building conventional CAD schemes. Various state-of-art segmentations can be applied to medical images. These techniques can be broadly categorized into (i) threshold-based (global, manual, adaptive, optimal, local adaptive, etc.), (ii) edge-based (Canny, Sobel, Fuzzy etc.), (iii) region-based, (iv) level-set based, (v) clustering-based, and (vi) artificial neural network-based methods. Depending upon the medical images we are dealing with and the ROI we need to segment, careful selection of the segmentation technique and finetuning to adapt to our needs is essential to achieve successful segmentation. Additionally, given the complexity of some medical images, it is not possible to develop a perfect segmentation algorithm. So, the ability to adjust to these automatically generated segmentation results must be integrated within the CAD scheme based on the user's expertise in reading and interpreting these images.

A few challenges that affect the output of segmentation are explained briefly in this section. First, when the ROI is not a single solid region and diffused into the background region, it results in a partial segmentation result. We need a bounding contour enclosing these regions to successfully identify all these diffused regions. Second, artifacts (noise) and intensity

inhomogeneity (shading artifacts) are common during image acquisition. Specific filtering and registration techniques are to be implemented to address these challenges. Third, soft tissues with closeness in gray level between foreground and background. This will cause leakages in the segmentation and can be addressed by defining a bounding box or stopping criteria during the segmentation growth.

2.4.2 Hypothesis and Proposed Approach

I have implemented several novel segmentation algorithms to address the segmentation challenges in my research studies. In the first study related to segmenting subtle soft tumor regions depicted on the CEDM images, I have designed an approach that uses the radiologists' center marking to begin the segmentation. Then, I a custom MLT region growing algorithm that performs two initial growth iterations using predetermined and adaptively adjusted thresholds. Then for the following repeated iterations, two key factors, including the growth rate and center shift, are monitored to decide the region's growth and stopping criteria. The segmentation results in this study are satisfactory and aided in building a robust CAD scheme for classifying breast lesions.

Second, in my studies [56] and [57], I implemented an iterative mapping-based adaptive thresholding to segment the intracranial brain region from the background, including surrounding artifacts with similar gray levels and diffused brain regions. The proposed CAD segmentation scheme first identifies three image markers that control the flow of the segmentation algorithm. During this mapping-based segmentation, the finetuning of segmentation begins at the global maximum image marker and is continued towards the other two image markers. This segmentation method uses the prior slice segmentation result to act as a reference for limiting boundary to avoid segmentation leakages and identify the multiple connected brain regions if existing. The segmentation results from both the CTP and CT brain images are accurate and need no manual

intervention to adjust the segmentation markings. Additionally, in the CTP study, the hypothesis is to quantify the blood flow over time by quantifying the asymmetrical blood volumes between the left and right hemispheres. For this purpose, I implemented a protocol that automatically performs various aspects of the CAD pipeline, including detection and segmentation of blood volumes to compute image markers to predict the prognosis of AIS patients. Similarly, in the aSAH study, the hypothesis is to quantify various brain regions, including developing image markers. I implemented adaptive multi-level thresholding followed by the mapped-based segmentation above to segment and quantify image markers to predict various clinical complications in aSAH patients.

Third, intracerebral hemorrhage (ICH) is the most common type of intracranial bleeding with stroke. The volume of ICH plays a critical role in calculating the ICH score, which is a well-known predictor of ICH prognosis. Currently, a qualitative approximation is performed using the ABC/2 method to compute the hemorrhagic volumes. In this study, my hypothesis is that building a more accurate, fast, and reliable interactive-CAD (ICAD) software tool to quantify the ICH volumes is possible and valuable for patient prognosis and suggests appropriate treatment methods. Thus, in this study, I developed a semi-automated segmentation ICAD tool for the quantitative estimation of ICH volume. The ICAD segmentation scheme uses the initial reference of the bounding box on a single slice enclosing the ICH volumes. Then, the adaptive thresholding will identify the true positive regions (ICH) on the current slice. These segmentation results will then be used to detect and continue the segmentation of ICH volumes in all the consecutive image slices until all the ICH blood is segmented. The results of this study were later validated by two experienced medical residents and have shown very encouraging dice similarity coefficient (DSC) ratings.

Additionally, I have developed various other semi-automated ICAD tools related to brain CT and MRI images [52], [58]–[60]. Clinical researchers used these tools to compute radiographic

image markers at our collaborating institutions to predict prognosis and observe an association with the clinical biomarkers of the various brain-related diseases such as stroke, aSAH, and ICH.

2.5 Bridging Gap Between Traditional Radiomics and Deep Learning-based CAD Schemes

2.5.1 Background

Traditional feature engineering to capture radiomic image markers is a popular and well-accepted methodology in developing CAD schemes of medical images. The radiomics approach captures underlying phenomena of suspicious masses by generating a vast number of study-relevant handcrafted features. These radiomic features can be obtained from a wide range of techniques, including preprocessing, segmentation, shape, density, texture patterns, frequency domain features etc. However, precautionary steps must be taken while extracting these features and identifying a small subset of useful features (avoid overfitting). Additionally, the radiomics approach also faces challenges in accurately segmenting subtle lesions in complex backgrounds and removing artifacts.

In contrast, convolutional neural networks (CNN) based CAD models avoid all the limitations of the traditional radiomic approach. In other words, CNNs avoid all segmentation, feature computation, and optimization to automatically generate higher-level features based on the objective of the task. Given a relatively large dataset, careful design, and finetuning, CNN-based CAD models can achieve comparative or even better results than traditional models. When the size of the medical dataset is relatively small, it is impossible to train the CNNs from scratch. In such circumstances, CNN's genetic features (i.e., shape, edge, general characteristics) learned from the natural images can be transferred as initial values and then later finetuned to suit the medical imaging task. However, the biggest limitations of CNNs in the medical image analysis are their

‘black box’ type, image-in, and prediction-out schemes, which are not easy to gain the trust of medical professionals.

2.5.2 Hypothesis and Proposed Approach

Many CAD schemes using either traditional or CNNs are developed and investigated separately to assess many medical challenges on relatively small image datasets. Thus, it is difficult to compare and evaluate these two schemes' performance and/or similarities. Thus, this study examines our hypothesis of observing the performance and correlation between the two schemes on a relatively large dataset. Therefore, this study explores the association/correlation between traditional radiomics feature-based CADs and deep learning framework-based CAD schemes classifying malignant and benign breast lesions. We also investigate whether the integration of these two types of features further improves performance in lesion classification.

During the traditional CAD pipeline design, many careful steps were considered to optimize each aspect of the design. These steps include (i) removal of chest wall, (ii) segmentation of lesions, (iii) extracting a wide range of radiomic features, and (iv) feature optimization and model designing. Additionally, while building the CNN model, we adapted the pre-trained ResNet architecture and finetuned and optimized it to fit our needs. Next, various ML models were then built to assess both the individual and combined performance of these two CAD schemes. The study results indicate that the CNN model yields significantly higher classification performance than the traditional radiomics model. It also supports our hypothesis that both radiomics and automated features contain highly correlated information in lesion classification. The details of this study will be presented in chapter 7.

3 Classification of Breast Masses Using A Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms

3.1 Introduction

Full-field digital mammography (FFDM) and dynamic contrast-enhanced breast magnetic resonance imaging (DCE-MRI) are two commonly used imaging modalities in breast cancer detection, diagnosis, and prognosis assessment. Mammography is the first line or the most popular breast imaging modality due to its high image resolution, improved image contrast, low operation cost, faster imaging scan and widely accessibility. However, as a two-dimensional projection imaging modality, mammography has relatively lower sensitivity and specificity due to the overlap of dense and heterogeneous fibro-glandular tissues (FGT) over the suspicious breast lesions. For example, mammography has lower sensitivity among women who are younger[61], have dense breasts[62], use hormone replacement therapy[63], and carry certain breast cancer susceptibility genes[64]. One study reported that mammography sensitivity reduced from 87.0% in women with almost entirely fatty breasts to 62.9% in women with extremely dense breasts or reduced from 83.3% in women over 80 years old to 68.6% in women younger than 50 years old[65]. Specificity of mammography is also low. During a 10-year screening period, more than 50% women would receive at least one false-positive recall and 7 to 9% have at least one false-positive biopsy[66], which adds anxiety with potentially long-term psychosocial consequences to many women[67].

On the other hand, DCE-MRI has superior sensitivity in detecting and/or diagnosing invasive breast cancer comparing to mammography and other existing breast imaging modalities[68]. However, DCE-MRI has a number of disadvantages including higher cost and longer imaging scanning time. It also has a relatively lower specificity, which may generate many unnecessary

breast biopsies and/or over-diagnosis[69]. As a result, both FFDM and DCE-MRI imaging modalities have advantages and disadvantage used in breast cancer detection and diagnosis.

In order to maximally take advantages of both FFDM and DCE-MRI imaging modalities, while overcome or reduce their disadvantages, an alternative imaging modality namely, contrast-enhanced digital mammography (CEDM) emerges and is quickly gaining momentum in the clinical trials worldwide. When using CEDM imaging modality, contrast agent is injected into breast and two series of scans are conducted at two different X-ray energy levels. Since malignant lesions are often accompanied by increased blood vessels that have unique permeability as compared to benign and/or normal tissues, the use of contrast agent allows analyzing morphology and vascular enhancement of the suspicious lesions. Additionally, when logarithmic subtraction is performed between two scans taken at different instances after injection of contrast agent, the difference in permeability is further enhanced and overlapping effect of FGT is removed. In general, CEDM generates multiple images including low energy (LE) images (similar to FFDM) and a contrasted enhanced dual-energy subtraction (DES) images (similar to MRI, but it is ~4 times faster than MRI exam). Therefore, CEDM has emerged as a promising imaging modality to overcome effect of tissue overlapping in FFDM and enable detection of tumor's neovascularity related functional information similar to MRI, while maintaining high image resolution as FFDM[70].

In reading and interpreting breast images, accurate classification between malignant and benign lesions is still a major challenge in current clinical practice. Studies have shown that performance of breast lesion diagnosis varied due to the intra- and inter-reader variability[71], and approximately only one in four biopsies are proved to be malignant[72]. Thus, in order to help improve accuracy and reproducibility in classification between malignant and benign breast

lesions, developing computer-aided diagnosis (CAD) schemes aiming to assist radiologists in their decision-making to better assess risk of malignancy of the detected suspicious breast lesions have been attracted extensive research interest in medical imaging field for the last two decades[73], [74]. Although CEDM is an emerging imaging modality, our recent pilot study has demonstrated that classification results based on a machine learning classifier that fuses the computed quantitative image features from CEDM images might provide complementary information to radiologists in particular to help reduce false-positives[75]. Thus, based on the well-developed CAD concept, objective of this study is to investigate a new and optimal approach to develop a fully automated CAD scheme of CEDM images and yield optimal performance in classification between the malignant and benign mass-type breast lesions.

3.2 Materials and Methods

3.2.1 CEDM Imaging Acquisition and Dataset

CEDM images used in this study were retrospectively collected from the existing clinical database of Mayo Clinic Arizona. All CEDM imaging examinations were performed using the following imaging acquisition protocol. In brief, the patient with mammography suite is seated to minimize vasovagal episodes and the intravenous line is first flushed with 10mL of saline. Next, an iodinated contrast agent of 1.5 mL/Kg of OMNIPAQUE 350 (GE Healthcare, Princeton, NJ, USA) is injected using a single lumen power injector at a rate of 3 mL/second. Last, the intravenous line is flushed again with an additional 10 mL of saline. If possible, the injected arm is raised above patient's head to facilitate contrast drainage from the arm, which enables maximum contrast circulation. After 2 minutes of contrast agent injection, the breast is compressed, and image acquisition starts.

In one CEDM imaging procedure, two sequential images on mediolateral oblique (MLO) and craniocaudal (CC) view are taken at both low and high X-ray energy levels. The low-energy (LE) image is acquired at (26-32kVp), which is less than the K-edge of iodine (33.2keV) to yield higher image contrast of soft tissue and calcifications similar to the regular FFDM. The high-energy (HE) image is acquired at an energy significantly higher than K-edge of iodine at (45-49kVp). Figure 3-1(a) shows the workflow for the CEDM imaging acquisition with approximate timestamps at each instance (view and energy). Finally, a difference (third) image is obtained by taking subtraction between HE and LE image, which is named as dual-energy subtracted (DES) image as shown in Figure 3-1(b). DES image is a single contrast medium-enhanced image that improves the visual enhancement of neovascularity information in and around the tumors while suppresses or removes the normal breast parenchymal or fibro-glandular tissues in the background. Figure 3-2 shows several examples in our dataset where the lesions are almost invisible or undetectable in LE (or regular FFDM) images, but they are clearly visible in DES images with the highly distinguishable lesion boundary contour from the parenchymal background.

In summary, from the clinical database, we retrieved and assembled a fully anonymous CEDM image test dataset that involves 111 women underwent breast cancer diagnosis at Mayo Clinic Arizona. Each case depicts one detected suspicious breast mass, which had biopsy. Based on the histopathologic test results of the biopsy samples, 78 masses were confirmed to be malignant and 33 were benign. In this dataset, each mass was considered visible in both CC and MLO views of LE images.

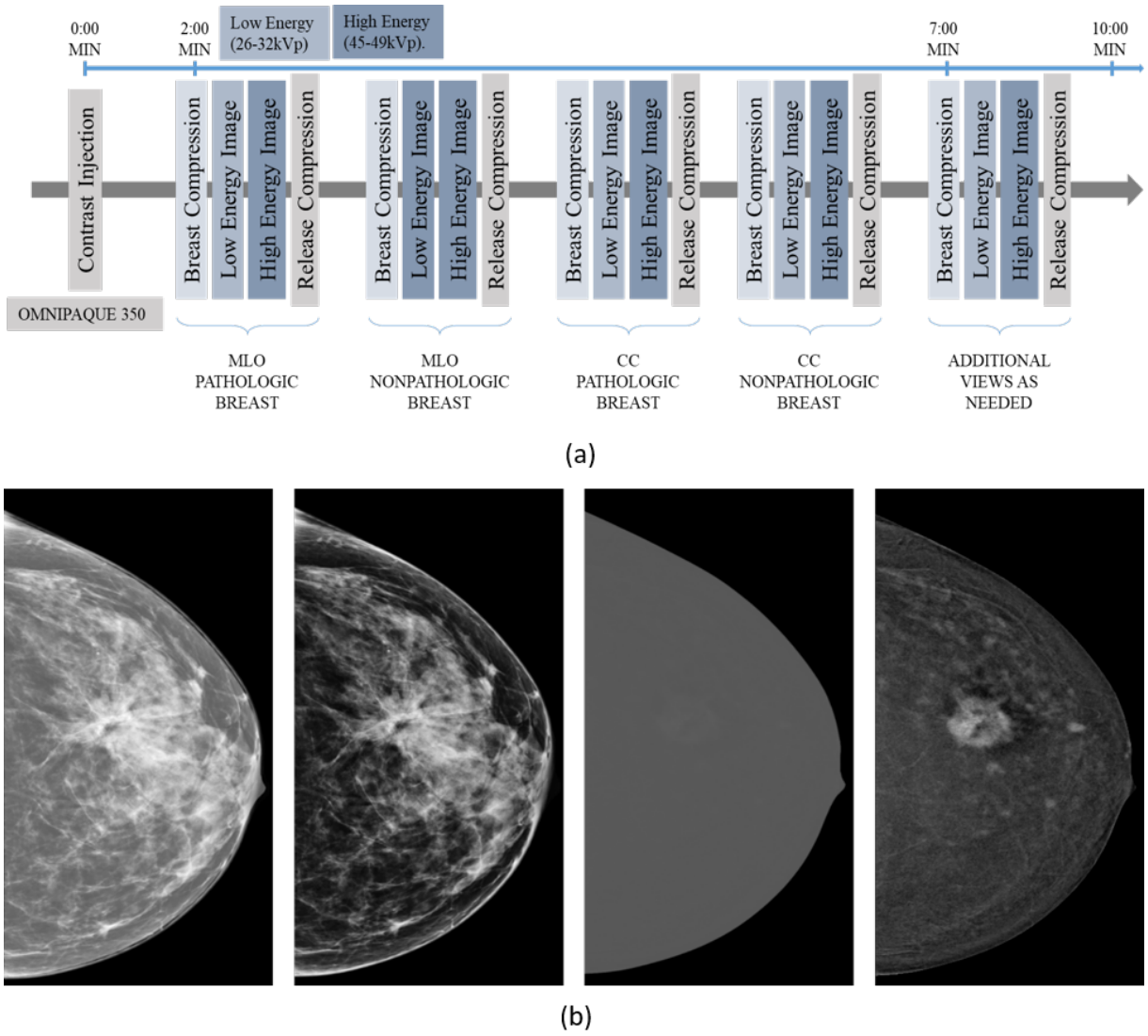


Figure 3-1: (a) Illustration of the workflow of a CEDM imaging acquisition procedure and (b) an example of 4 images from left to right: High energy (HE) image, Low energy (LE) image, dual energy subtraction (DES) image displayed at same window and level as HE image, and the DES image displayed at an adjusted window and level for improving visibility, respectively.

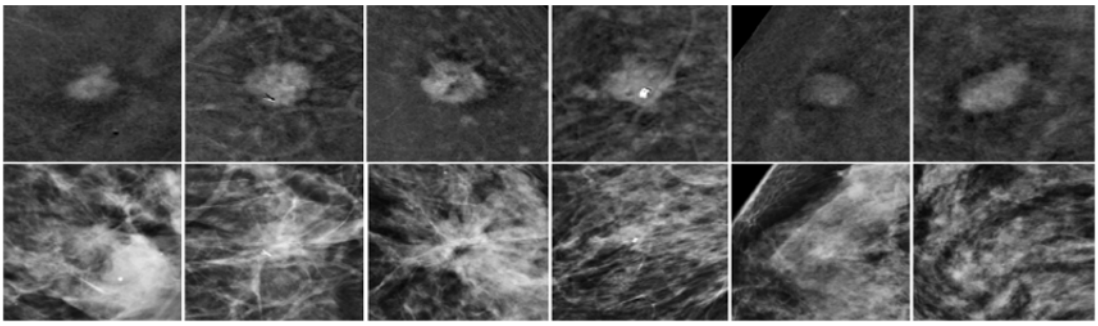


Figure 3-2: A few samples in which mass-type lesions are clearly visible in DES images (the 1st row), but almost invisible in LE (or regular FFDM) images (the 2nd row).

Similar to the regular FFDM images, size of the original images acquired from CEDM is either 3328 × 2560 or 4096 × 3328 pixels depending on breast size. Then, based on the standardized approach to develop a CAD scheme for detecting and/or classifying breast masses[76], the original images were subsampled using bilinear interpolation method in which output pixel value is a weighted average pixel value from a 2-by-2 neighborhood kernel. The subsampled image size was reduced to corresponding 666 × 512 or 820 × 666 pixels. Similar image subsampling process has been commonly used in developing previous CAD schemes of FFDM images (e.g., [77]). Thus, this image dataset was used to develop and assess a new CAD scheme to classify between malignant and benign masses.

3.2.2 Breast Mass Segmentation

The first step of our CAD scheme is to automatically segment suspicious mass region depicting on each image of interest. Since CEDM is a diagnostic imaging modality and it applies to the recalled patients who have suspicious lesions detected in the screening mammograms, the locations of each suspicious mass-type lesion in two CC and MLO view images are already known and can be easily mapped to CEDM images. Figure 3-3 shows the graphical user interface (GUI) of our new interactive CAD scheme of CEDM images. After loading an image (either CC or MLO view) of interest in the GUI, the user can observe and place an initial seed point around the mass

center to segment the mass region in future testing cases. In this study, all region growth seeds namely, the mass region center pixels, were automatically placed based on the clinical truth file. In a batched CAD processing, no human intervention is involved. Although a large number of mass segmentation methods or algorithms have been developed and reported in the literature[78], we in this study applied and implemented a multi-layer topographic (MLT) region growing segmentation algorithm, which has been well-developed and applied in a number of previous CAD schemes[16], [79].

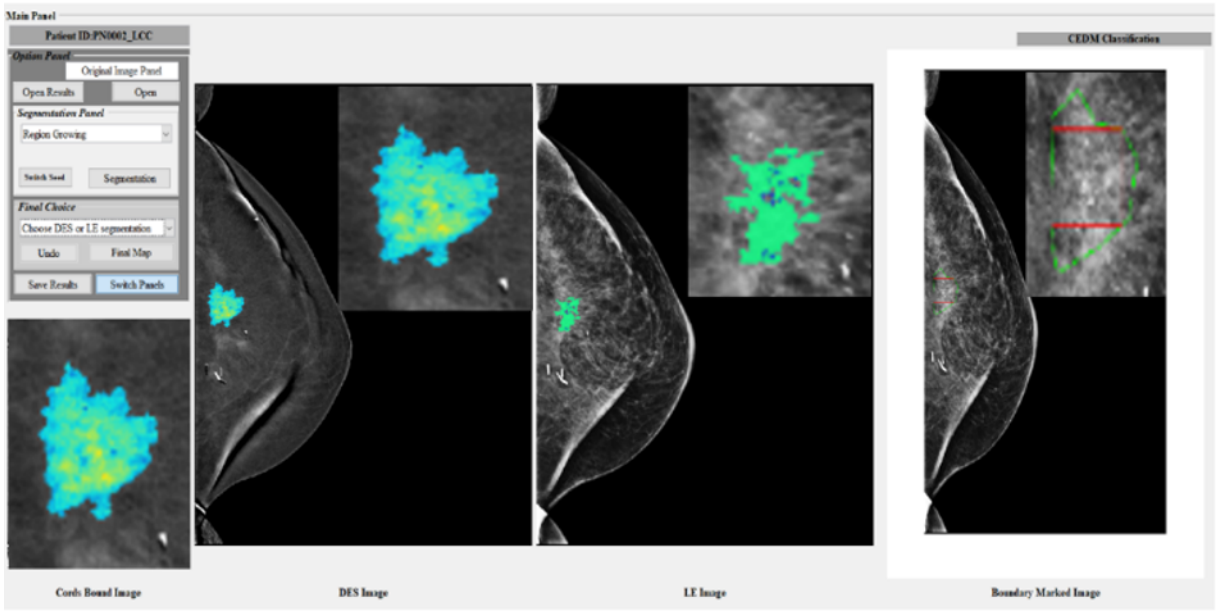


Figure 3-3: Illustration of graphical user interface (GUI) of the CAD scheme.

In brief, the MLT region growing algorithm first applies with a conventional region growing process using a pre-selected small threshold to segment the central region of the lesion. Second, the threshold value is adaptively adjusted based on the pixel intensity difference between the initially segmented region and the surrounding region. The next layer of segmentation is performed with the adjusted threshold. Two parameters namely, growth rate (an increase of size) and center shift (the displacement of centroid) between the prior and current segmentation or region growth

layer, are computed. If the current growth layer passes two boundary conditions in which the growth rate is less than 100% (double the size), and the shift of the region center is less than 10 pixels, this current growth layer is accepted to replace its prior growth layer (region). Third, this region growing process continues to define the new layer until the new growth layer fails to pass one of the above two boundary conditions. Then, the growing iteration ends, and the last "prior" growth layer is selected as the final segmentation output. Figure 3-3 shows examples of the mass segmentation results on both DES and LE images (from the left to right). For a comparison, image with radiologist's marking on the mass region is also displayed in the first image from the right.

3.2.3 Feature Computation

After segmentation of each breast mass, the second step of CAD is to compute image features. In the development phase, CAD initially computes a set of 109 image features, which can be divided into 4 groups as listed in Table 3-1. The first group includes 4 mass size and shape related image features, which include mass size, the maximum radius or convexity (smoothness) of mass boundary. The second group includes 13 statistical features related to heterogeneity of mass density (pixel values). The third group includes 8 features to detect variation of density (pixel values) between the mass and its surrounding boundary. These features have been defined and used in our previous CAD schemes of different types of medical images (including FFDM images and lung CT images) to represent segmented lesions[80], [81].

Table 3-1: Summary of 109 features separated by feature class.

Feature class	Feature number	Feature description
Shape	4 (F1-F4)	Mass size, convexity, maximum radius, standard deviation (STD) of all radii.
Tumor Density related	13 (F5-F17)	Mean, STD, energy, entropy, maximum intensity, mean absolute deviation, median, minimum, range, rms, uniformity, skewness, and kurtosis of a segmented mass region.
Tumor background Density related	8 (F18-F25)	gradient mean, gradient STD, ISO-intensity, fluctuation mean, fluctuation STD, mean contrast, contrast, STD ratio of mass to boundary.
Wavelet	84 (F26-F109)	Apply the density features on the four wavelet components

Last, the fourth group includes 84 wavelet transform generated image features. Specifically, a two-dimensional wavelet transform (using a "Coiflet 1" filter) was applied, which decomposes each image into four decompositions. During the decomposition, two-dimensional filters (low pass and high pass) are applied in both x- and y-direction to compute I_{LL}, I_{LH}, I_{HL}, and I_{HH} as represented in Figure 3-4. For instance, I_{HL} is obtained by applying a high pass filter along the x-direction followed by a low pass filter in the y-direction as described in Equation 3.1, where L and H indicate low and high pass filters, respectively. N_H and N_L are the length of filters for high and low pass filter, respectively. In our study both N_H and N_L have length of 6. All features in the second and third groups are applied individually on each of the four wavelet components to detect density variations in the filtered wavelet decompositions.

$$I_{HL}(i,j) = \sum_{p=1}^{N_H} \sum_{q=1}^{N_L} H(p)L(q)I(i+p, j+q) \tag{3.1}$$

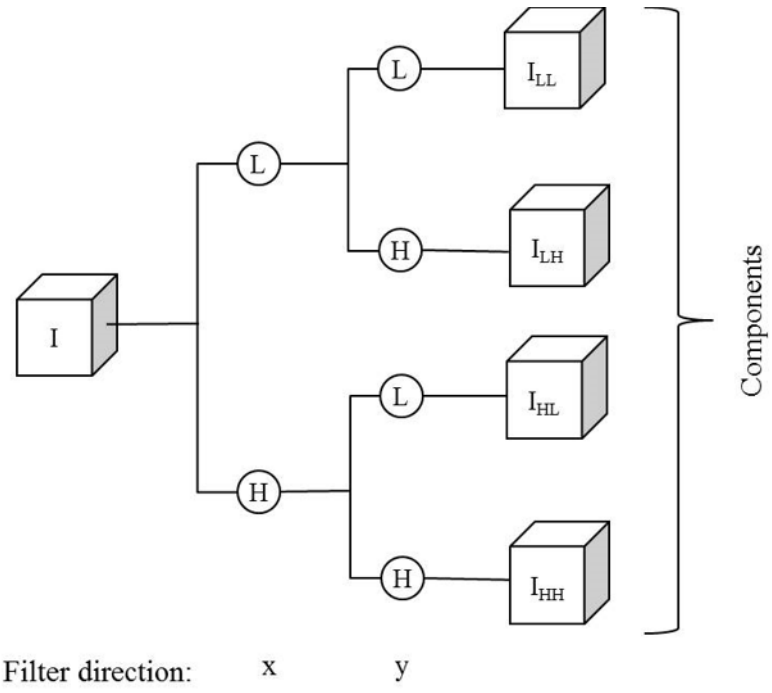


Figure 3-4: Illustration of the image decomposition using a wavelet transformation (one-level, un-decimated two-dimensional wavelet transforms using "Coiflet 1" filter), where L is a low pass filter and H is a high pass filter.

For non-solid or diffused breast lesions, since there are multiple suspicious masses spread in the images without any connectivity between them, the segmented primary (the largest) mass region is used for computing shape, morphology, and background related features, whereas all the pixels in the diffused suspicious masses are used for calculating density related image features, which are independent of its corresponding background information.

In addition, we took two considerations in CAD feature computation. First, each mass is segmented separately from CC and MLO view images. Two segmented mass regions from two view images often do not have the exact same computed image feature values due to the different tissue overlapping in two 2D projection images. Thus, we used average value of two feature values separately computed from either CC or MLO view image to represent the final feature value of a mass of interest. Second, due to the possible difference of mass region segmentation results on LE and DES images, GUI of our CAD scheme has a function that allows user to select an optimal

segmentation result from either LE or DES image, and then map the selected segmentation result to the matched DES or LE images if necessary, in the future clinical applications. Using this mapping method, we are able to compute optimal image features from both LE and DES images.

3.2.4 Machine Learning Classifier and Performance Assessment

The third step of CAD uses a multi-feature fusion-based machine learning classifier to produce a classification score for each suspicious mass under test, which ranges from 0 to 1. The higher classification score represents a higher likelihood of the region being malignant. Although many machine learning classifiers have been used in developing CAD schemes, in this study, we selected a simple and popular machine learning classifier namely, a multilayer perceptron (MLP) based artificial neural network to classify suspicious breast mass. For this purpose, we used Weka data mining and machine learning software platform[82] to train and test the MLP classifier. In order to build a highly performed and robust machine learning classifier, we needed to consider and address following 3 issues: (1) a relatively small CEDM image dataset of 111 cases, (2) a relatively large pool of initially computed 109 features, and (3) imbalance of the cases in the CEDM dataset, which includes 29.7% (33/111) of benign masses and 70.3% (78/111) of malignant masses.

To overcome or minimize the potentially biased impact of above 3 issues, we adopted following 3 methods. First, we applied a leave-one-case-out (LOCO) cross-validation method to maximize learning power while minimizing the case partition and testing bias[83]. Second, we used a correlation-based feature subset (CFS) evaluator to reduce the dimensionality of feature space by dropping highly correlated, redundant, irrelevant and noisy features, and thus produce a subset of optimal features from the initial feature pool[84]. Specifically, a CFS evaluator was set to use a BestFirst search method with a search termination setting of 5, which means if the number of non-improving nodes in the forward search is greater than 5, CFS stops feature selection

process. The features selected before the termination were used to build an optimal feature set to train the classifier. Third, we applied a Synthetic Minority Oversampling Technique (SMOTE)[85] method to generate synthetic data of benign masses to produce a more balanced training dataset to avoid or minimize the potential classification bias towards majority (malignant) cases. For example, we applied SMOTE to double "benign cases" from 33 to 66. Thus, the dataset becomes more balanced with 45.9% (66/144) benign cases and 54.1% (78/144) malignant cases. The effectiveness of applying similar SMOTE method has been applied and tested in our previous studies[86], [87].

After taking these considerations and protection steps, we built 4 MLP classifiers. The first 2 MLPs used image features computed from the segmented mass regions depicting on either DES or LE images, respectively. Since mass segmentation results on DES and LE images may vary significantly. Using the GUI tool of our CAD scheme (as shown in Figure 3), we mapped the better or optimal segmentation results from one image to another (i.e., from DES to LE or vice versa). Then, after optimal mapping, CAD recomputed image features from the mapped mass regions depicting on either LE or DES images.

In the process of training and testing each MLP classifier, we embedded both feature selection (a CFS evaluator) and SMOTE algorithm into the LOCO cross-validation process. Specifically, in each LOCO training and testing iteration, one mass region was first removed from the training dataset. Second, SMOTE algorithm was applied to generate synthetic data to double the number of benign cases. Third, a CFS feature selection evaluator was applied to select a set of optimal features. Last, a MLP classifier was trained using the training dataset and selected optimal features. After training process, the classifier was applied to test one independent testing mass, which was not involved in the training process. This CFS evaluator and SMOTE algorithm embedded LOCO

training and testing iteration process was repeated 111 times. As a result, each of 111 masses in our dataset was independently tested. The classification scores of all 111 cases were thus generated and recorded.

Finally, classification performance of each MLP classifier was evaluated using following two methods and evaluation indices. First, a receiver operating characteristic (ROC) method was used. Each ROC curve and the area under ROC curve (AUC) were computed using a maximum likelihood based ROC curve fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/, University of Chicago). Second, we applied an operating threshold (T = 0.5) on the classification scores to divide original 111 masses into two classes (or groups) of malignant and benign cases. From the results, we generated a confusion matrix and computed overall classification accuracy, as well as the positive and negative predictive values (PPV and NPV). The evaluation results of 4 MLP classifiers were then tabulated and compared.

3.3 Results

Figures 3-5 to 3-7 show examples of comparing the results of applying our CAD scheme to segment regions of the same breast masses depicting on both DES (the 1st row) and LE (the 2nd row) images, respectively. The results show that due to the large heterogeneity of breast masses and surrounding parenchymal tissue background, results of mass segmentation vary between using LE and DES images as compared to the regions of interest (ROIs) marked by the radiologists (as shown in the third row of Figure 3-5 to Figure 3-7). In general, for masses that are partially occulted under the surrounding dense fibro-glandular tissues, it is often difficult for CAD to generate satisfactory segmentation results using LE images due to the great fuzziness of mass boundary. For illustration purpose, Figure 3-5 shows 6 examples in which segmentation failed in LE images (the middle row) as compared to the better segmentation results yielded using DES images (the

top row). On the other hand, some masses may be invisible or only partially visible on DES images due to the lack of enhancement or large necrosis. In these cases, CAD segmentation results on LE images may more accurately represent the real mass regions (see Figure 3-6). Figure 3-7 shows examples of the mapped "optimal" segmentation results on both LE and DES images.

The 3rd row of Figures 3-5 to 3-7 also shows the lesion bounding boxes placed by radiologists. By comparing with CAD-generated segmentation results (as shown in the 1st and 2nd rows of these figures), we can clearly observe that CAD-segmented lesion boundary are often more accurate than the results of manually drawing.

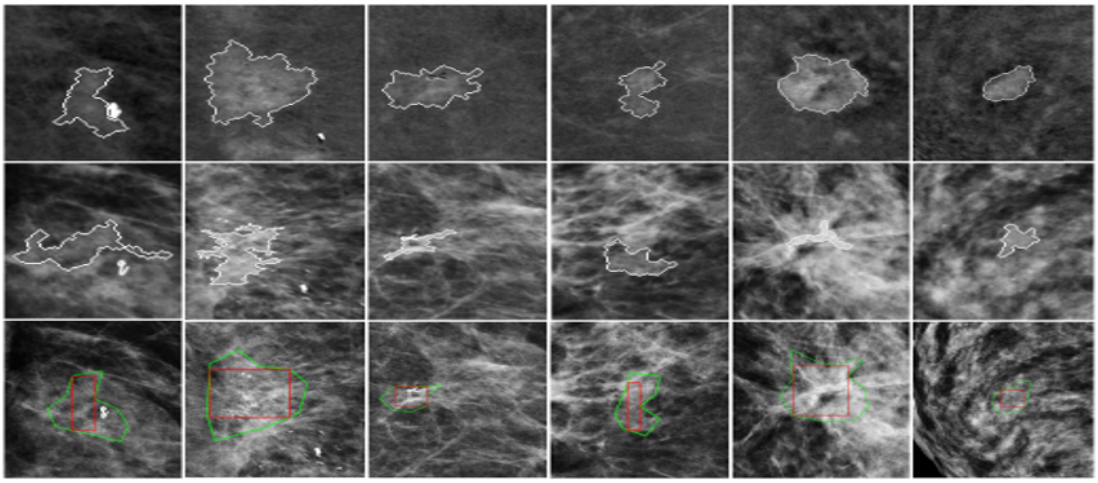


Figure 3-5: Sample cases illustrating failed segmentation in LE images (2nd row) as compared to DES images (1st row). The 3rd-row shows the lesion bounding boxes placed by the radiologists.

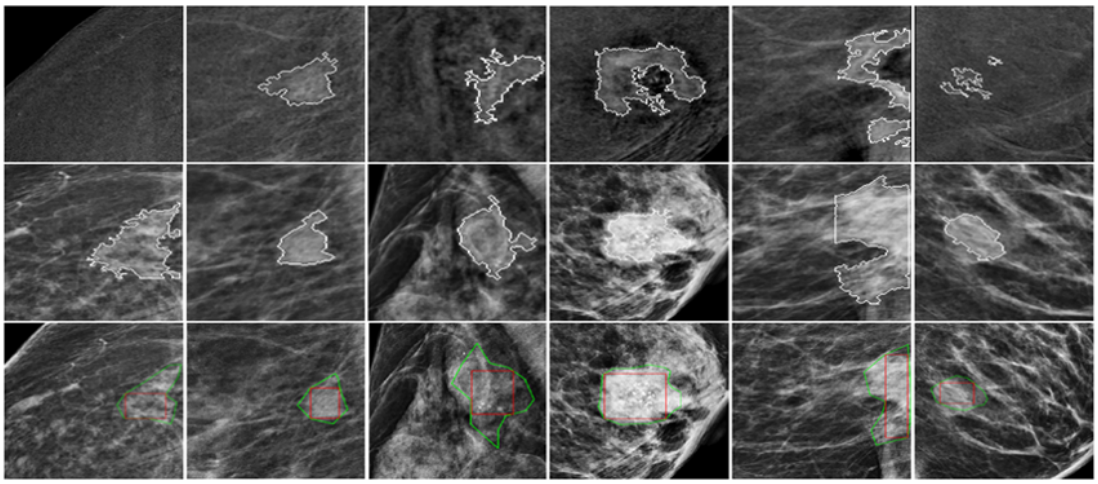


Figure 3-6: Sample cases illustrating failed segmentation in DES images (1st row) as compared to LE images (2nd row). The 3rd-row shows the lesion bounding boxes placed by the radiologists.

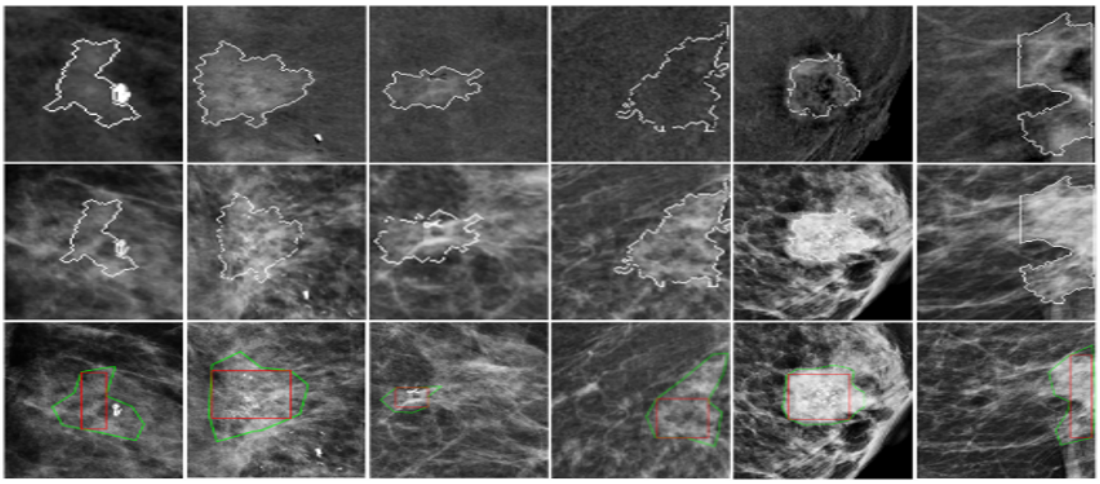


Figure 3-7: Sample cases showing optimal segmentation mapping on both DES (1st row) and LE (2nd row) images. The 3rd-row shows the lesion bounding boxes placed by the radiologists.

Table 3-2 lists the highly performed image features, which were selected more than 90% of LOCO cross-validation based 144 training iterations. From this Table, a number of interesting observations can be made. For example, (1) although lesion shape or boundary margin features (i.e., F1 to F4 as shown in Table 3-1) are commonly considered as the most important image features in many of previous CAD schemes, this type of features were largely removed or not selected by the classifiers trained using LE images, which indicates that the lesion boundary features can only play important role when the lesions are more accurately segmented. (2) The density heterogeneity features computed from both inside a lesion and its surrounding background

can contribute to the CAD scheme to classify between malignant and benign lesions. (3) Extracting optimal density heterogeneity features can also expand to the filtered images (i.e., using wavelet transform as done in this study). From the filtered images, CAD can detect and select optimal features to build the machine leaning classifiers.

Table 3-2: List of features selected in ≥ 90% of LOCO training and testing of 4 MLP classifiers.

Original segmentation				**Optimally mapped segmentation**			
DES images		**LE images**		**DES images**		**LE images**	
Feature	%	Feature	%	Feature	%	Feature	%
F1	97	F7	99	F2	100	F5	100
F2	99	F10	99	F5	98	F6	100
F3	100	F11	98	F8	95	F10	100
F5	100	F16	100	F12	100	F16	100
F20	100	F21	93	F20	100	F33	100
F41	99	F24	100	F35	100	F41	100
F48	95	F39	100	F41	100	F43	97
F56	91	F44	100	F50	97	F44	100
F65	100	F45	100	F62	100	F45	100
F66	100	F65	100	F77	99	F46	99
F83	100	F80	100	F81	97	F65	92
F88	100	F81	94	F82	99	F75	100
F101	100			F83	99	F88	100
				F88	100	F103	100
				F96	100		
				F109	100		

Figure 3-8 plots and compares 4 ROC curves that are generated using 4 sets of CAD classification scores, which were computed by 4 MLP classifiers. Since in this dataset, 7 masses were not enhanced in CEDM images (i.e., one mass region as shown in the first ROI of the top row of Figure 6) and thus they cannot be segmented, the first MLP classifier was trained and tested using the remaining 104 cases (27 benign vs. 77 malignant masses). Other 3 MLP classifiers were trained and tested using all 111 masses (33 benign vs. 78 malignant masses). The computed AUC values are 0.759±0.053 and 0.753±0.047 for the first two MLP classifiers, which were trained and tested using the mass regions originally segmented from DES and LE images, respectively. By mapping better or optimal segmentation results from LE images to DES images, AUC = 0.739±0.048, which did not show improvement of classification performance. However, when mapping the better segmentation results from DES images to LE images, AUC value of using the new MLP classifier significantly increases to 0.848±0.038 as compared to all other 3 MLP classifiers (with $p < 0.01$).

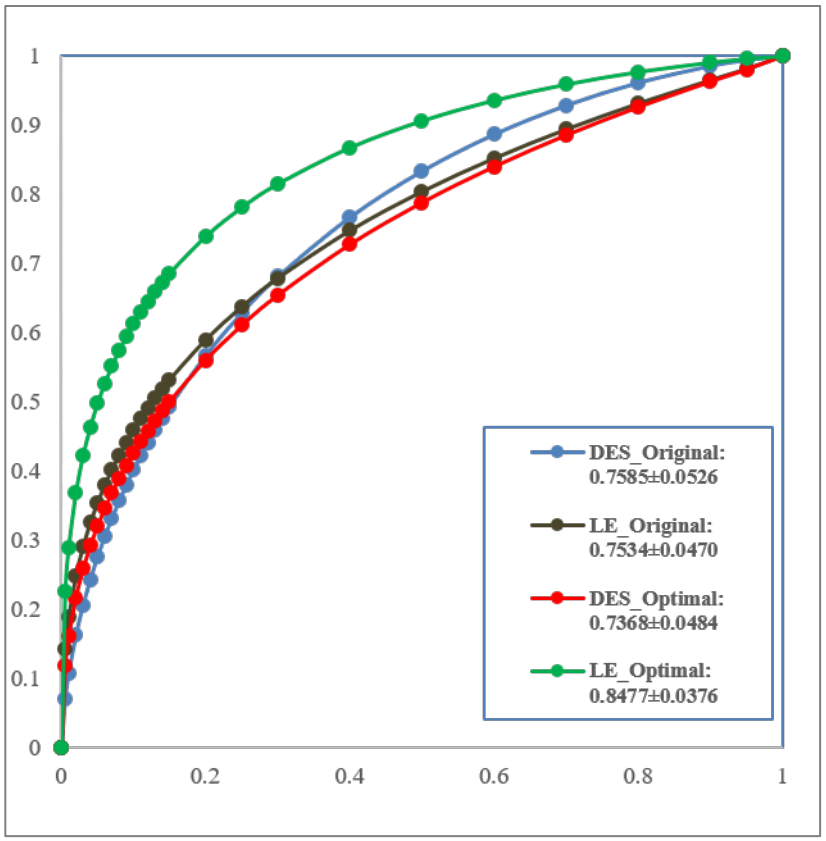


Figure 3-8: Comparison of four ROC curves generated using 4 MLP classifiers using the original and optimally mapped mass segmentation results on DES and LE images to distinguish between malignant and benign breast masses.

By applying an operation threshold of 0.5 on the MLP-generated classification scores, Table 3-3 and Table 3-4 shows two sets of 4 confusion matrices that were generated based on the distribution of the classification scores of 4 MLP classifiers applying to DES and LE images twice. Two confusion matrices of Table 3-3 show distribution of the classification scores computed by two MLP classifiers trained using the originally segmented mass regions, while two matrices in Table 3-4 show distribution of the classification scores computed by two MLP classifiers trained using the optimally mapped mass regions depicting on DES and LE images, respectively.

Table 3-3: Two confusion matrices generated when applying MLP classifier to the originally segmented breast mass regions depicting on LE and DES images.

Actual / Prediction	LE Images (Total: 111 masses)		DES Images (Total: 104 masses)	
	Malignant	Benign	Malignant	Benign
Malignant	56 (0.5)	9 (0.08)	53 (0.51)	9 (0.09)
Benign	22 (0.2)	24 (0.22)	24 (0.23)	18 (0.17)

Table 3-4: Two confusion matrices generated when applying MLP classifier to the optimally mapped segmented breast mass regions depicting on LE and DES images.

Actual / Prediction	LE Images (Total: 111 masses)		DES Images (Total: 111 masses)	
	Malignant	Benign	Malignant	Benign
Malignant	63 (0.57)	9 (0.08)	55 (0.49)	12 (0.11)
Benign	15 (0.13)	24 (0.22)	23 (0.21)	21 (0.19)

From these 4 confusion matrices, the overall classification accuracy, positive predictive values (PPV) and negative predictive values (NPV) of 4 MLP classifiers were computed and compared as shown in Table 3-5. The results indicated that using the fourth MLP classifier, which was trained and tested using LE images after mapping the optimal mass region segmentation results from DES images to LE images, yielded the highest classification accuracy, as well as the highest PPV and NPV values. For example, when comparing to the second MLP classifier trained and tested using

the originally segmented mass regions depicting on LE images, the overall classification accuracy of the fourth MLP increased 8.7% (from 72.1% to 78.4%).

Table 3-5: Summarization and comparison of classification performance using 4 MLP classifiers.

MLP Classifier	Original segmentation		Optimally mapped segmentation	
	DES images	LE images	DES images	LE images
Overall accuracy	68.3%	72.1%	68.5%	78.4%
AUC±STD	0.769±0.053	0.753±0.047	0.737±0.048	0.848±0.038
PPV	85.5%	86.2%	82.1%	87.5%
NPV	42.9%	52.2%	47.7%	61.5%

3.4 Discussion

Our recent study supports the concept of applying a CAD scheme or method to assist radiologists for interpreting CEDM images in classifying between malignant and benign lesions[75]. Thus, this is our first study to investigate how to optimally develop a fully-automated CAD scheme of CEDM images to classify between malignant and benign breast masses. This study has following unique characteristics and/or observations. First, in breast cancer detection and diagnosis, accurate classification between malignant and benign breast lesions remains a challenging task to date. Although CAD schemes of FFDM and breast MRI images have been developed aiming to assist radiologists in classifying between malignant and benign breast lesions in a large number of previous studies, these CAD schemes have not been accepted and used in the clinical practice. One of the primary difficulties is the lack of capability of accurately segmenting breast lesions depicting on images, in particular, the FFDM images due to the fuzzy lesion boundary caused by tissue overlapping. Segmentation of breast lesion is not only difficult for CAD, but also for radiologists, which generates large intra- and inter-reader variability. Thus,

inaccurate lesion segmentation reduces the accuracy and robustness of computing image features that are used to develop machine learning classifiers. In CEDM imaging modality, DES images enable to enhance breast lesion regions, while removing and/or suppressing normal parenchymal tissues that overlap or surround the lesions. Thus, segmentation of breast lesion regions from DES images becomes much more accurate and robust. This is a unique contribution of including DES images in the CAD schemes. This study demonstrated that by mapping the optimal lesion segmentation results on DES images to LE images, our CAD scheme or MLP classifier yielded significantly higher performance in mass classification than using the CAD scheme applying to the originally segmented mass regions depicting on LE images.

Second, although using DES images enhances lesion boundary and makes lesion segmentation easier and more accurate than using LE images, it also has potential disadvantages in developing CAD schemes. For example, we observed that after contrast enhancement, lesions depicting on DES images become more homogeneous, which lose much density heterogeneity information of the lesions depicting on LE images. Thus, when using density heterogeneity and texture related image features computed from the segmented lesions to train and develop machine learning classifiers, CAD classification performance using DES images does not yield significantly higher performance than using LE images. It seems that the advantage of more accurate lesion segmentation using DES images is partially cancelled out by its disadvantage of losing density or texture heterogeneity information. As a result, if we want to improve CAD classification results using the lesion regions segmented from DES images, different strategy and/or image features need to be explored and used in future studies.

Third, since CAD performance heavily depends on the difficult and diverse levels of datasets, it is not feasible to directly compare lesion classification performance (i.e., AUC = 0.848±0.038)

of CAD scheme developed in this study for CEDM images with other previously developed CAD schemes of FFDM images, which reported AUC values ranging from 0.70 to 0.87 due to use of different datasets[88]. However, despite the limitation of a relatively small and unbalanced dataset with 111 cases (33 benign vs. 78 malignant cases), this study is valid because we have taken 3 measures namely, (1) a leave-one-case-out (LOCO) cross-validation method, (2) a correlation-based feature subset (CFS) evaluator-based feature selection method and (3) a synthetic minority oversampling technique (SMOTE) method. Both CFS and SMOTE were embedded into LOCO cross-validation. In order to support advantage of this embedded approach, we also tested CAD performance by removing SMOTE and CFS. Table 3-6 compares the performance changes. We observed that (1) when SMOTE was not applied to balance the dataset (33 benign, 78 malignant), the performance reduced as comparing to the embedded method used in this study, and (2) when the CFS feature selection step was also removed, the performance further decreased.

Table 3-6: Comparison of classification performance changes between three methods.

Method	Original segmentation				Optimally mapped segmentation			
	DES images		LE images		DES images		LE images	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Proposed MLP	0.76±0.05	68.3%	0.75±0.05	72.1%	0.74±0.05	68.5%	0.85±0.04	78.4%
Remove SMOTE	0.63±0.06	70.19%	0.70±0.05	68.46%	0.75±0.05	62.16%	0.69±0.05	65.76%
Remove CFS and SMOTE	0.56±0.07	65.38%	0.71±0.05	66.67%	0.61±0.06	64.86%	0.59±0.06	63.06%

Fourth, besides a MLP classifier, we have also applied the same CFS evaluator and SMOTE algorithm embedded LOCO training and testing iteration method to build a number of several other popular machine learning models or classifiers, which include logistic regression (LR), Bayesian belief network (BNN), k-nearest neighbor (KNN), Random Forest (RF) and Random Committee (RC) algorithms, which are available in Weka data mining software platform[82], to

classify between malignant and benign masses using DES and LE images. Although the performance levels of different classifiers vary (i.e., from the lowest AUC = 0.735±0.047 for logistic regression to the highest AUC = 0.895±0.030 for BNN when using LE images after mapping the optimal lesion segmentation results from DES images), the performance change trend in each classifier maintains consistent. This supports the results produced using the MLP classifier as reported in the Results section of this paper. The additional testing results using different types of machine learning classifiers clearly indicate when using the original lesion segmentation, classification performance levels on DES and LE images are quite comparable. However, when mapping the optimal lesion segmentation results generated on DES images to LE images, all classifiers using different machine learning models yielded the highest classification performance.

Table 3-7: Comparison of CAD performance between the new averaging (CC, MLO) method applied in this study and the conventional independent CC and MLO method.

Method	Original segmentation				Optimally mapped segmentation			
	DES images		LE images		DES images		LE images	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
Average (CC, ML)	0.76±0.05	68.3%	0.75±0.05	72.1%	0.74±0.05	68.5%	0.85±0.04	78.4%
Independent (CC, ML)	0.73±0.04	63.28%	0.81±0.03	74.77%	0.72±0.04	63.96%	0.78±0.03	72.07%

Fifth, unlike the most of CAD schemes in digital mammograms, which classify lesion based on the image features computed from one (i.e., either CC or MLO) view image, we fused image features computed from two mass regions depicting on two view images into one classifier. In order to demonstrate the advantages of this new fusion approach, we did a comparison experiment. Table 3-7 shows the results to compare this new and conventional CAD approach, which

demonstrate that using the averaging features in developing the CAD schemes yields the higher performance.

Last, his study also has a number of limitations. First, the size of dataset remains small. Thus, the performance and robustness of the CAD scheme of CEDM images need to be further optimized and validated using new large and diverse image dataset in the future studies. Second, in the proof-of-concept study to demonstrate the feasibility of developing a fully-automated CAD scheme of CEDM images, we used well-developed CAD pipeline with new lesion segmentation mapping methods and image features mainly focusing on density heterogeneity of a lesion and its surrounding background. Thus, more studies in developing potentially new CAD approaches may also need in future studies.

In summary, we investigated and tested a new approach to develop a first automated CAD scheme of breast lesion classification using CEDM images. Study results demonstrated that LE and DES images generated from CEDM contain complementarily valuable information. Using DES images may help more accurately segment suspicious lesions from the images if the lesions are enhanced. Then, by mapping the optimal lesion segmentation results (lesion boundary contour) from DES images onto LE images, the density heterogeneity and texture-based image features can be more accurately computed from LE images. Therefore, the lesion classification performance of using this new CAD scheme that combines these two types of images can be significantly improved. In summary, new knowledge that we learned from this study helped establish a solid foundation for us and/or other researchers in the CAD field to continue developing and optimizing novel CAD schemes of CEDM images with improved performance in the future studies.

4 Developing New Quantitative Ct Image Markers to Predict Prognosis of Acute Ischemic Stroke Patients

4.1 Introduction

Stroke is the fifth leading cause of death with more than 140,000 deaths each year in USA [89]. Ischemic stroke is the most common stroke accounting for about 87% of all stroke cases. It occurs when a vessel supplying blood to brain is obstructed due to the narrow blood vessel or clogs with fatty deposits. In particular, the acute ischemic stroke (AIS) due to large vessel occlusion (LVO) poses a large cerebral tissue at risk and carries high morbidity and mortality of the patients [90]. Leading cause of LVO is cardio-embolism due to atrial fibrillation. There has been an increased incidence of atrial fibrillation with aging population that parallels the increase in LVO related AIS [91]. Besides applying clot-dissolving drug and tissue plasminogen activator to treat AIS patients in attempt to restore blood flow in the blocked brain regions, endovascular mechanical thrombectomy (EMT) is also recommended to treat some AIS patients with LVO to reduce morbidity and mortality [92]. Recently, several multicenter clinical trials have reported improved treatment outcome in AIS patients with LVO [93], [45]. In a study of operating EMT 6 to 24 hours after stroke with a mismatch between deficit and infarct, the results showed the improved outcome in patients treated with EMT plus standard care as compared to patients who received standard care alone [93]. Additionally, similar improvement was reported in another study conducted EMT in the stroke patients at 6 to 16 hours [45]. These trials are based on the estimation of cerebral infarct core and existence of salvageable radiological brain tissue “at-risk” for infarction. Case selection biases in these trials are also widely debated. Additionally, the objective clinical or radiological correlation for region-specific AIS related outcome after such interventions is still unknown and needs an investigation in further studies.

Due to the lack of accurate clinical markers to accurately stratify AIS patients who can or cannot benefit from EMT to date [94], identifying new clinical markers that highly associate with efficacy of EMT plays a critical role to quickly restore the peripheral blood supply in a short time period, which can help minimize the amount of brain tissue injury or risk of permanent tissue damage among the AIS patients who can benefit from EMT. Principles of currently used imaging software to select patients for EMT depends on arterial input function (AIF) and venous output function (VOF) to provide estimate regarding cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT) and time to peak (TTP or Tmax) computed from computed tomography perfusion (CTP) images [46]. Such algorithms that depend on initial contrast flow in a major cerebral arterial system and outflow through a major venous channel are unable to capture microcirculatory dynamics of contrast flow through the brain parenchyma. Using CTP derived CBV to estimate ischemic core or AIS volume is often inaccurate as comparing to that estimated using magnetic resonance imaging (MRI) based diffusion weighted imaging (DWI) [47]. Some researchers highlighted that these inconsistencies are partially contributed due to high noise sensitivity in deconvolution-based on singular value decomposition. However, even though newer techniques using Bayesian method are robust as compared to SVD, they are still missing a significant infarct volume visible on the MRI images. In a recent study, our group showed that clinical outcome highly associated with the final cerebral infarct volume estimated using DWI sequences of post-intervention MRI [60]. Hence, although CTP has significantly clinical advantages over MRI in AIS diagnosis due to its highly efficient and wide accessibility, it requires developing new imaging markers that have significantly increased prediction accuracy or high association to the clinical outcome of AIS patients with LVO.

Despite of the improved imaging technology, qualitative image assessment of AIS severity or status using the radiologists-rated cerebral blood flow, cerebral blood volume, mean transit time and time to peak [95] has limitations including lack of quantitative assessment and large inter-reader variability [48]. In order to overcome these limitations, identifying and developing new quantitative image markers have been attracting broad research interests in the medical image informatics field [96]. In particular, the recently developed radiomics concept has proven that radiographic images (i.e., CT and MRI) depict useful image phenotype features that highly associate with genomic biomarkers [97] and have potential to predict disease prognosis [98]. Thus, based on radiomics concept and previous research focus in developing computer-aided detection and diagnosis (CAD) schemes of medical images [15], [99]–[102] including a CAD scheme equipped with an interactive graphic user interface (GUI) to detect and quantify severity of aneurysmal subarachnoid hemorrhage patients using brain CT images [103], we aim to investigate the feasibility of developing new quantitative image markers computed from CTP images at an early diagnosis stage to predict AIS prognosis in this study. For this purpose, we developed a new CAD scheme with several novel image processing algorithms to segment the contrast agent enhanced blood volumes in bilateral cerebral hemispheres of brain, generate cumulative blood flow curves and then compute asymmetrical blood flow related features in two brain hemispheres. Then, image markers based on the best single feature and machine learning (ML) models fused with multi-features are developed and tested to predict clinical benefit or outcome in group of AIS patients undergoing EMT for LVO. The details of the study design and experimental data analysis results are presented in the following sections of this article.

4.2 Materials and Methods

4.2.1 Image Dataset

A de-identified retrospective dataset of pre-intervention CTP images of 31 AIS patients due to LVO was obtained from the Department of Neurology at the University of Oklahoma Health Sciences Center (OUHSC). Based on current clinical standard of Modified Rankin Scale (mRS) [104], the primary treatment outcomes of AIS patients are categorized into seven scales (from 0 to 6). Figure 4-1(a) summarizes the distribution of mRS among these 31 patients. Due to the small dataset, we divided the patients into two classes of good (favorable) and poor (unfavorable) prognosis based on mRS as shown in Figure 4-1(b). Class-0 includes 16 cases in which mRS range from 0 – 3 representing from no symptoms to moderate neurological disability (requiring some help, but able to walk unassisted), while calss-1 includes 15 cases ranked from 4 – 6 in mRS representing from moderately severe disability (unable to walk and attend to bodily needs without assistance) to dead. The goal of this study was to develop and apply new quantitative image markers to classify AIS cases into these two classes.

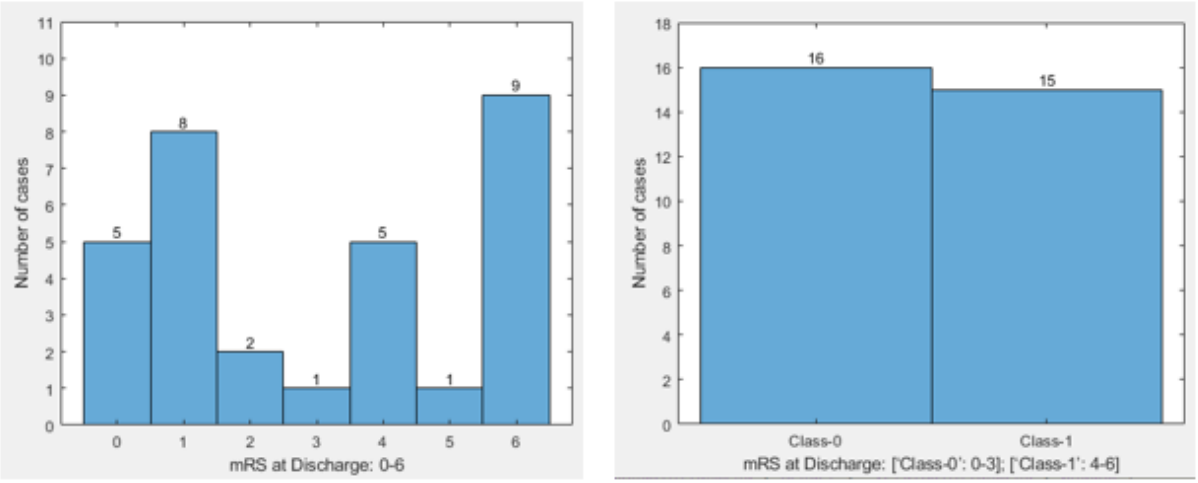


Figure 4-1: Distribution of patients based on the Modified Rankin Scale (mRS). (a) Separated by individual mRS, (b) Separated by mRS into two classes: ['class-0': 0-3]; ['class-1': 4-6].

During AIS diagnosis and treatment in OUHSC, each patient is pre-assessed radiologically for their eligibility to undergo EMT. Specifically, during the image acquisition, multiple CT image scans are conducted including initial non-contrast CT of the head, CT angiogram, and sequential CTP scans, respectively. In CTP image acquisition, a rapid intravenous infusion of 40 ml of Isovue-370 contrast agent is administered. This contrast agent gradually enters and passes through the bloodstream and vessels, which helps visually distinguish blood flow from other brain structures. Thus, during the CTP image acquisition process, the dynamic flow of contrast agent (i.e., wash-in and wash-out patterns) are used to capture the total blood amount and velocity of blood flow through different regions of the brain.

4.2.2 Image Pre-processing

Among the clinical cases, there is variability in the number of CTP scanning sequences (i.e., ranging from 28 to 89), scanning range (i.e., whole brain or only the targeted volume of interest) resulting in different number of image slices in one scanning sequence (i.e., ranging from 8 to 23), and image slice thickness (i.e., 2.5 or 5.0mm). The pixel spacing parameters for all the cases are 0.488 × 0.488 mm for length and width respectively. In addition, some cases can use one-directional CTP scanning protocol, while others use two-directional scanning protocol. In one-directional scan, when completing one scanning sequence of the targeted brain section, CT machine pauses and returns to the starting point to perform the next scanning sequence (i.e., always from top to bottom or vice versa). Whereas in use of two-directional protocol, CT machine cyclically captures image sequences continuously without any break (i.e., scanning from top to bottom and then reversely scanning from bottom to top). Regarding image reconstruction algorithm, we used GE CT scanners, automatic settings of filtered back projection (FBP) with a standard convolution kernel and the filter selected as head. Then, the image pixel intensity values

are converted into standard Hounsfield Unit (HU) value based on the dicom rescaling parameters such as slope and intercept.

In order to accommodate the acquired CTP image sequence irregularities mentioned above and make all cases comparable to each other, it requires CAD scheme to automatically organize the images retrieved from the clinical picture archiving and communication system (PACS) database by identifying the number of CTP image scanning sequences and adaptively labeling each image slice to a specific indexed brain location in the correct image sequence. Since the head is held fixed during image acquisition, the degree of similarity between the matched brain sections during the adjacent scanning sequences is higher. Thus, CAD scheme uses a simple dice-similarity based approach to identify two parameters (scan-type namely, one-directional or two-directional scanning protocols, number of the detected unique brain matching sections in different scanning sequences) for each case.

In our dataset, we find that the maximum number of scanned image slices in one scanning sequence for all patients is ≤ 23 images. Thus, for each case, CAD scheme selects the first 50 images retrieved from each CTP case to initially identify scan-type and match images in two adjacent scanning sequence. For this purpose, CAD scheme first performs a rough brain segmentation using thresholding to exclude the skull region. If the number of connected regions in an image slice is more than one, only the most significant area is included, while others are discarded. Next, CAD scheme uses the dice similarity coefficient to compute the degree of similarity among these segmented areas in the images to identify the first unique CTP scanning sequence and its best matching pair. This can determine the scan-type and the actual number of CTP image slices in each scanning sequence. Figure 2 illustrates this CAD process. For example, for one case using one-directional scanning protocol with total slice number of x, CAD scheme

requires to detect its scan-type and the number of x slices in one scanning sequence. Thus, slice (1) matches slice (x+1) and continues (as shown in the middle row of Figure 4-2). The bottom row of Figure 2 shows that if the case uses two-directional scanning protocol with total slice number of x+2 in one scanning sequence, CAD will identify image matching in different order (i.e., slice x+2 matches slice x+3, and so forth). Once detecting the scan-type and the number of image slices in one sequence using the initial set of 50 image slices, CAD maps the results to the rest of all images in one case. Thus, the total number of CTP image scanning sequences of each case is computed using Equation (4.1) and all images of every series are labeled to a specific indexed brain location.

$$Total\ number\ of\ CTP\ sequences = \frac{Total\ number\ of\ images\ in\ one\ case}{\text{number of image slices in one CTP sequence}} \tag{4.1}$$

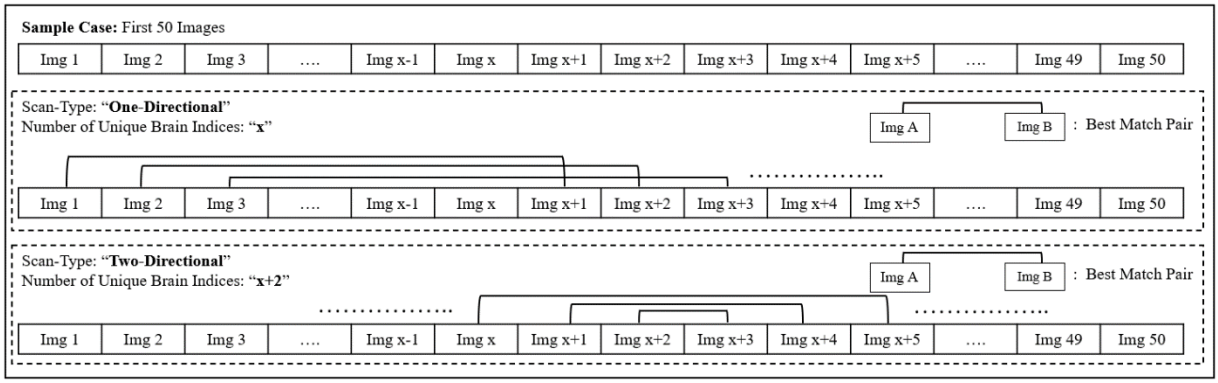


Figure 4-2: A sample illustration of the proposed dice-similarity based approach identifying the parameters (scan-type and number of unique brain indices).

4.2.3 Image Segmentation

A novel segmentation algorithm using image markers or fiducials and mapping technique is implemented in CAD scheme. For each unique CTP image sequence, CAD scheme identifies three perfusion markers namely, global minimum (g_m), local minimum (l_m), and maximum peak (m_p). Like the method discussed in image pre-processing, CAD first performs an initial brain segmentation using thresholding to identify the largest connected intracranial brain region. A line

plot depicting the initially segmented areas of each slice is constructed. The slice with the largest area is considered as m_p, the slice with the smallest area is regarded as g_m, and the slice with the smallest area in the opposite direction concerning the m_p slice and g_m slice is marked as l_m. Next, the fine-tuning procedure is applied to segment final brain areas beginning with the m_p slice as the initial starting point because the actual brain in this slice is usually the most significant, single connected component, and enclosed within the skull. Therefore, a precise segmentation without any leakages can be attained on this slice using a thresholding-based segmentation. In this way, the fine-tuning segmentation process continues towards either left or right direction, applying a consecutive mapping technique. This segmentation method uses the prior slice segmentation result to act as both a reference for limiting boundary to avoid segmentation leakages and identifying the multiple connected brain regions if existing. The limiting boundary criterion prevents segmentation leakage, which can be corrected applying a process of morphological dilation. In the case of multiple connected regions, each region is examined and constrained to enclose within the limiting boundary for inclusion in the current slice segmentation. In summary, this segmentation process initiates at m_p slice for each sequence and continues using the steps as mentioned above until either a g_m slice or l_m slice is reached in both directions covering all the image slices.

Since this study primarily focuses on understanding and analyzing the asymmetry of blood flow between the left and right hemispheres of the brain, which is an important image feature used by neuro-radiologists to assess the efficacy of EMT, CAD scheme splits each segmented brain slice into two parts of left and right hemispheres. We also design and implement an interactive graphical user interface (GUI) of the CAD scheme with multiple visual-aid tools and functionalities as shown in Figure 4-3. Additionally, if GUI shows a slight tilt in brain image orientation occurred during image acquisition, a function tool has been added in the GUI to request

CAD scheme to rotate images and correct image orientation. Thus, CAD scheme can correctly separate left and right hemispheres of the brain for image feature computation and data analysis.

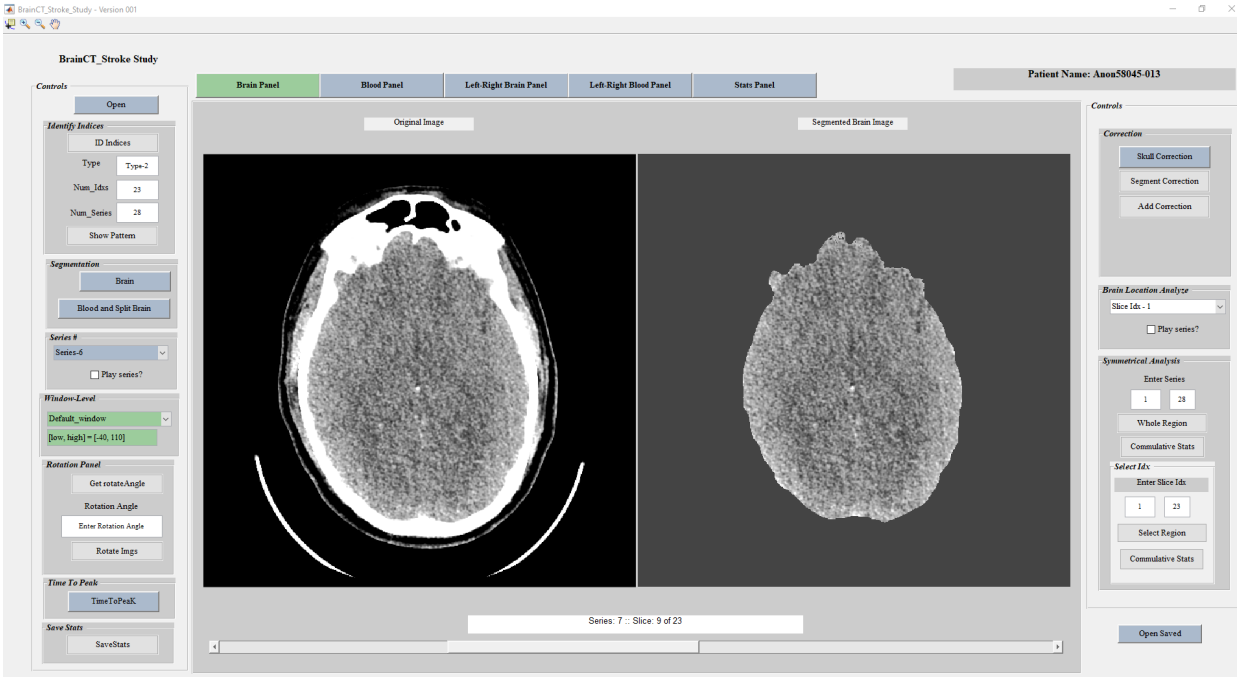


Figure 4-3: Picture of the implemented interactive graphical user interface (GUI) of the CAD scheme, which includes two image windows showing the original CT image slice (left) and the segmented brain area (right), and multiple operating functionalities and parameter assignment boxes on both left and right column.

4.2.4 Image Feature Computation

CAD scheme first computes blood volume in the left and right hemispheres from blood profile image of each CTP slice. After grouping the computed blood values per each unique sequence, CAD computes the following image features. First, the blood volume (V_{Blood}) in one CTP slice is computed using Equation (4.2),

$$V_{Blood}(slice) = N_{B\cdot P}\,(\text{slice}) \times P_L \times P_B \times S_T\,(\text{slice}) \tag{4.2}$$

where $N_{B.P}$ is the number of detected or segmented blood pixels, P_L and P_B represent pixel length and breadth, while S_T represents slice thickness, respectively. A summation of V_{Blood} for all slices for one CTP scanning sequence (or a series) is computed using Equation (4.3),

$$V_{Blood}(series) = \sum_{slice_k=1}^{m} \left(N_{B.P}(\text{slice}_k) \times P_L \times P_B \times S_T(\text{slice}_k)\right) \quad (4.3)$$

where m = number of slices in the series. However, V_{Blood} per series can be represented as two independent terms voxel count (VC_{term}), and VP_{term} as explained from Equations (4.4-4.6).

$$VC._{term} = N_{B.P}[slice_{k=1}, \quad slice_{k=2} \ \ldots \quad slice_{k=m}] \quad (4.4)$$

$$VP._{term} = P_L \times P_B \times S_T[slice_{k=1}, \quad slice_{k=2} \ \ldots \quad slice_{k=m}] \quad (4.5)$$

$$V_{Blood}(series) = VC._{term} \times (VP._{term})' \quad (4.6)$$

where VP_{term} is an array of constant values for all the series. VC_{term} changes for each series depending upon the blood profile. Thus, V_{Blood} is directly proportional to VC_{term}. Throughout the rest of this article, we represent the summation of the VC_{term} as V_{Blood}, as indicated in Equation (4.7).

$$V_{Blood}(series) \propto \left(VC._{term} = \sum_{slice_k=1}^{m} N_{B.P}(\text{slice}_k) \right) \quad (4.7)$$

Second, to detect the trend of blood supply in two brain hemispheres over time, CAD scheme computes the cumulative volume of blood ($V_{Cumulative}$) as shown in Equation (4.8),

$$V_{Cumulative} = \left[V_{s_1}, V_{s_1+s_2}, V_{s_1+s_2+s_3}, \ldots, V_{s_1+s_2+s_3+\cdots+s_{n-1}+s_n}\right] \quad (4.8)$$

where $V_{s_1+s_2+s_3+\cdots+s_{n-1}+s_n}$ is the summation of V_{Blood} between 1st to nth CTP scanning series, and n indicates the number of images in a unique series. This cumulative volume of blood ($V_{Cumulative}$) is calculated independently for both left ($V_{Cumulative\_L}$) and right ($V_{Cumulative\_R}$) hemispheres from their respective blood profile images using the steps as mentioned earlier from Equations (4.2-4.8).

Third, due to variation in operator settings during CTP image acquisition, the scan brain regions and duration of scanning vary among the patients. If a line plot is mapped between the number of unique series (x-axis: n) and cumulative volume of blood in a hemisphere (y-axis: [$V_{Cumulative\_L}$ or $V_{Cumulative\_R}$]) for all the cases, the scales will not be compatible. To address this, we performed a case-based normalization of n, $V_{Cumulative\_L}$, and $V_{Cumulative\_R,}$ as shown in Equations (4.9-4.11) to scale or normalize the computed feature values between 0 and 1.

$$\hat{n} = \frac{n - \min(n)}{\max(n) - \min(n)} \tag{4.9}$$

$$\mathrm{Min_{val}} = \min\left(\mathrm{V_{Cumulative_L}}, \mathrm{V_{Cumulative_R}}\right); \mathrm{Max_{val}} = \max\left(V_{Cumulative_L}, V_{Cumulative_R}\right) \tag{4.10}$$

$$\hat{\mathrm{v}}_{Cumulative\_L} = \frac{\mathrm{V_{Cumulative\_L}} - \mathrm{Min\_val}}{\mathrm{Max\_val} - \mathrm{Min\_val}}\ ;\ \hat{\mathrm{v}}_{Cumulative\_R} = \frac{\mathrm{V_{Cumulative\_R}} - \mathrm{Min\_val}}{\mathrm{Max\_val} - \mathrm{Min\_val}} \tag{4.11}$$

4.2.5 Asymmetrical Blood Flow Pattern Analysis

Since acute ischemic stroke (AIS) usually occurs in one hemisphere of the brain with LVO, which blocks the respective primary arterial blood flow. As a result, blood flow velocity in two hemispheres of the brain is different, which makes the dynamic flow (wash-in and wash-out of contrast agent) go faster in the healthy hemisphere without LVO than the diseased hemisphere with LVO. Thus, detecting and quantifying asymmetrical blood flow patterns is our focus to identify new image markers to predict AIS prognosis. Specifically, as shown in Figure 4, we divide the timed image sequences into three equal phases namely, initial phase, intermediate phase, and final phase. For each phase, the corresponding section of the line segments ($V_{Cumulative\_L}$ and $V_{Cumulative\_R}$) from the line plots is utilized to calculate intermediate slopes (m_{AB}) using a linear regression method, where A is right or left hemisphere, and B is the phase. For example, m_{L2} is the slope of the left hemisphere in the 2nd/intermediate phase. Additionally, we subtracted the values of $V_{Cumulative\_L}$ and $V_{Cumulative\_R}$ between the left and right hemispheres to construct an

absolute difference in cumulative volumes ($|V_{Cumulative\_D}|$) followed by normalization to yield $\hat{v}_{Cumulative\_D}$. Then, $\hat{v}_{Cumulative\_D}$ curve is also divided into three equal phases similar to the method discussed earlier to compute their respective intermediate difference slopes (m_{DB}), where D represents the difference line, and B is the phase (as shown in Figure 4-4). Finally, the absolute cumulative disparity in the total blood volume between both hemispheres at the completion of CTP image acquisition is computed using the last value of the arrays ($\hat{v}_{Cumulative\_R(end)}, and\ \hat{v}_{Cumulative\_L(end)}$) as shown in the Equation (4.12).

$$abs\_difference_{Total} = abs\left(\hat{v}_{Cumulative\_R(end)} - \hat{v}_{Cumulative\_L(end)}\right) \quad (4.12)$$

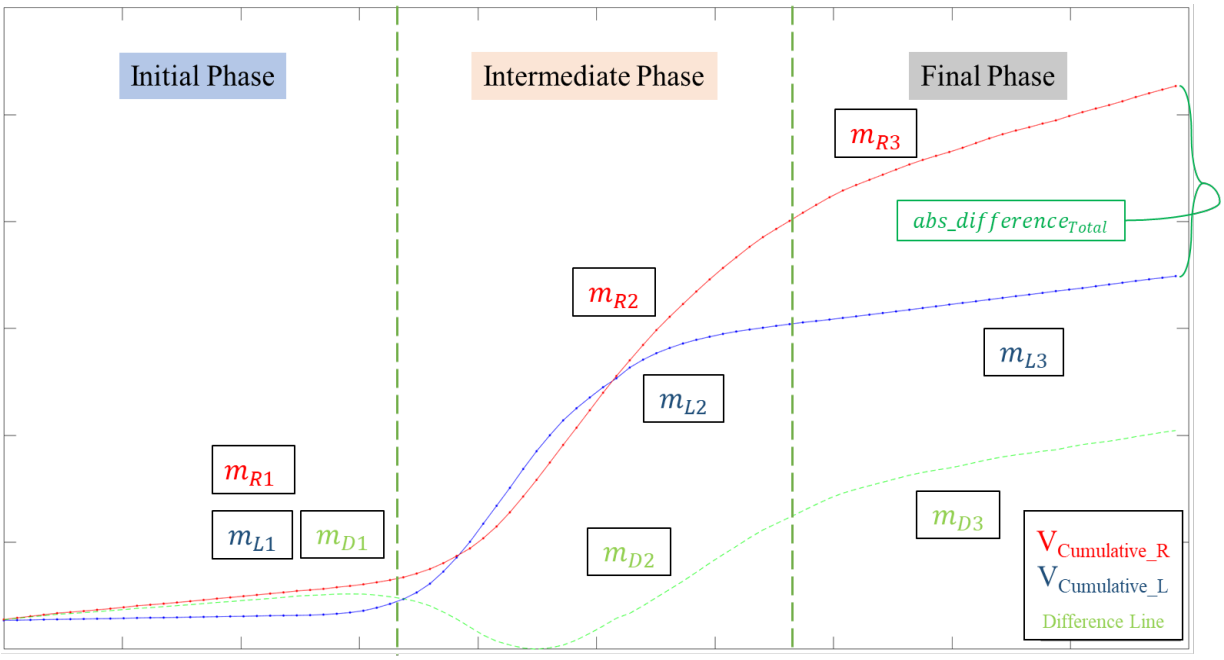


Figure 4-4: A sample illustration of sectoring cumulative volume of blood line plot into three equal phases and computing corresponding intermediate slopes for left and right hemisphere.

4.2.6 Machine Learning Classifier Model

Using the absolute cumulative disparity value computed at the end of the blood line plot ($abs\_difference_{Total}$) and other slop-based features computed in 3 phases (as shown in Figure 4), we test several models to develop image markers to classify cases into two mRS classes of

good and poor prognosis. Each model or marker generates classification scores ranging from 0 and 1. The higher score represents the higher likelihood of the case having poor prognosis ('class-1': mRS = 3-6). Table 4-1 lists 3 independent models.

1. Model-I only uses one image feature of the absolute cumulative disparity value ($abs\_difference_{Total}$) to simulate what neuro-radiologists do to predict patient prognosis with the quantitative data.

2. Model-II is built using features computed separately from two blood flow curves of the left and right hemispheres of the brain ($\hat{v}_{Cumulative\_L}, \hat{v}_{Cumulative\_R}$) in three phases and the difference in blood cumulative volume ($abs\_difference_{Total}$).
3. Model-III is built using features computed from one subtracted blood flow curve between the left and right hemispheres of the brain ($\hat{v}_{Cumulative\_D}$) in three phases and the differences in blood cumulative volume ($abs\_difference_{Total}$), which emphasizes on the bilateral asymmetry of the computed image features between left and right hemisphere of the brain.

Table 4-1: List of features included in each type of ML model.

Model	Feature Pool
Model-I	$abs\_difference_{Total}$
Model-II	$m_{L1}, m_{L2}, m_{L3}, m_{R1}, m_{R2}, m_{R3}, and\ abs\_difference_{Total}$
Model-III	$m_{D1}, m_{D2}, m_{D3}, and\ abs\_difference_{Total}$

In order to build multi-feature fusion models (Model-II and Model-III), we select two well-known supervised machine learning (ML) architectures namely, support vector machine (SVM)

and K-nearest neighborhood (KNN). Based on our previous experience of applying SVM and KNN in developing CAD schemes of medical images [105], [106], a polynomial kernel is used in SVM model and K = 5 (neighbors) is applied in KNN model. To build the optimal ML models, following three steps are used. First, a feature-wise normalization is performed to transform values of each feature to a scale from 0 to 1. Second, a principal component analysis (PCA) method is applied to generate new feature vector with a variance rate of 95% applied to reduce redundancy of the image features. Third, due to the small dataset, a leave-one-case-out (LOCO) based cross-validation method is adopted to train and evaluate each ML model to maximize the number of training samples and avoid case patrician bias [107]. In this way, each of 31 cases in our dataset will be independently tested by the model trained using other 30 cases in 31 training iterations.

To evaluate the performance of each classification model, we used the following two steps. First, a receiver operating characteristic curve (ROC) is constructed form the classification scores and the area under the ROC curve (AUC) is computed and used as an assessment index to evaluate and compare the performance of each model to classify between two mRS classes. Second, we apply an operating threshold on the classification scores (T = 0.5) to divide all testing cases into two mRS classes (score $\leq$ 0.5: 'Class-0'; score > 0.5: 'Class-1'). From the classification results, several confusion matrices corresponding to different models are generated, which are used to compute various performance indices (i.e., classification accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity).

In summary, Figure 4-5 shows a complete flow diagram of each step in our CAD scheme, which includes image pre-processing and data analysis pipelines using the ML model. All programs used in CAD scheme and GUI tool (as shown in Figure 4-3) are coded using MATLAB R2019b package and libraries. Whereas the ROC curve and the AUC are computed using a

maximum likelihood-based ROC curve fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/ , University of Chicago), which is publicly available and widely used in radiology and medical imaging informatics field.

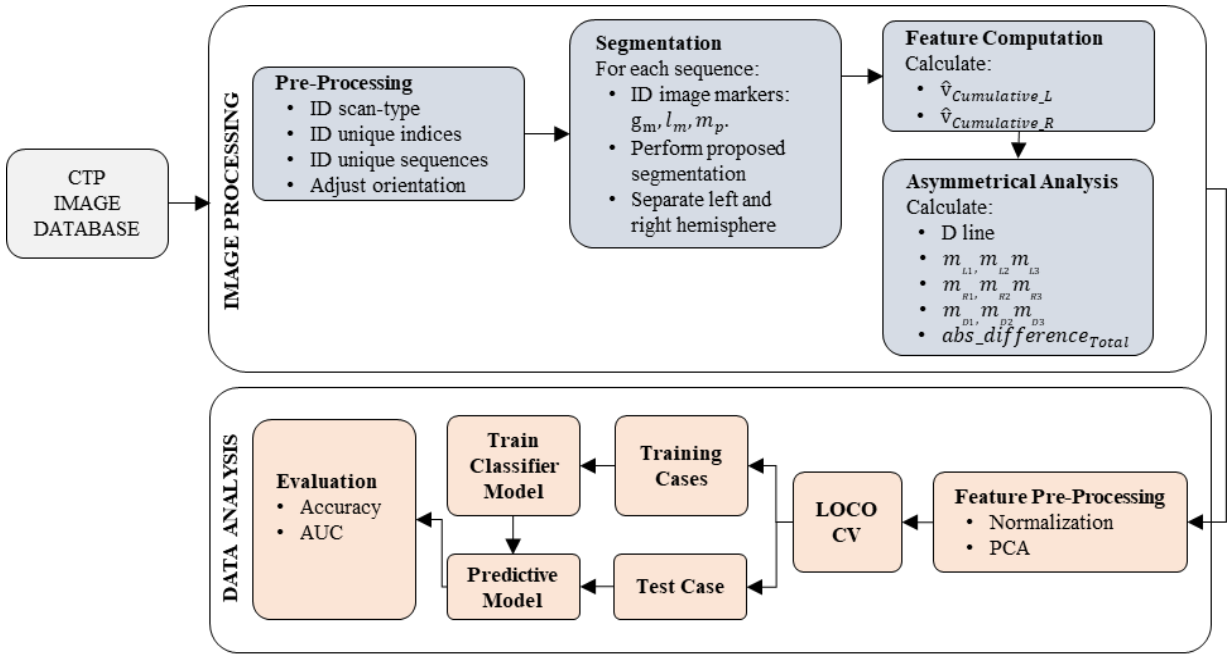


Figure 4-5: A detailed flow diagram of each step of the proposed CAD scheme.

4.3 Results

Figure 4-6 shows an example of the matched brain CTP slices in the whole scan of 28 sequences, which depict change in the visibility of the contrast agent in the blood over the CTP image acquisition time. Looking at these 28 images from the top left to the bottom right in a left to right fashion, one can notice that the visibility of blood contrast is more dominant in the right hemisphere at the early phase of CTP scans as compared to the left hemisphere (wash-in). As the CTP acquisition continues, the blood contrast in the right hemisphere appears to drain completely (wash-out) first, whereas the contrast is still visible in the left hemisphere in the images of later scanning sequences. This example shows that our CAD scheme automatically detects the amount

of contrast filled blood volume over time using the cumulative amount of blood to quantify the contrast agent variation between the left and right hemispheres of the brain.

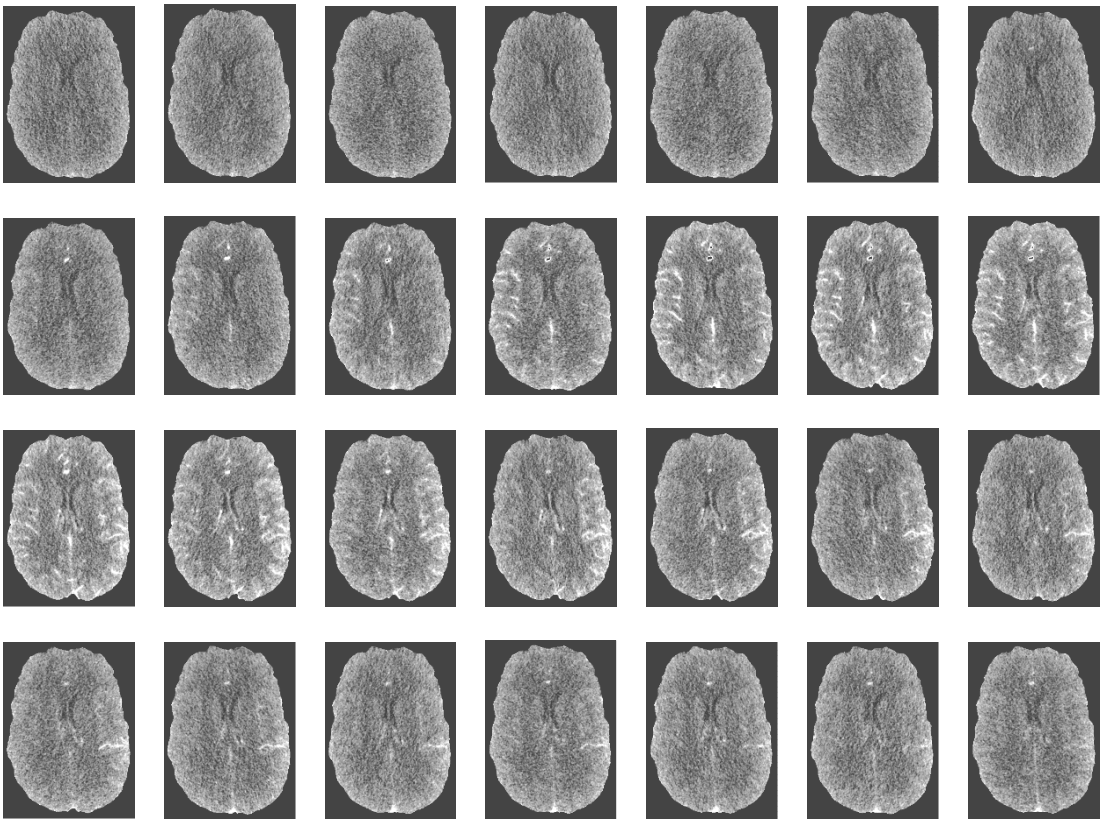


Figure 4-6: From top-left to bottom-right: A sample brain index over CTP acquisition time depicting the variation in blood flow between the left and right hemisphere.

Figure 4-7 demonstrates the segmentation results of a sample CTP brain series for an individual patient. Unlike a regular segmentation scheme that targets to identify either a single connected cerebral region or uses the skull as the limiting boundary may fail to achieve the accurate results as it may miss certain regions or cause possible leakages in some other scenarios, our CAD scheme successfully detects the multiple connected regions as shown in the 2nd and 3rd images of Figure 4-7. The reason of generating such a result is because our CAD scheme uses 3 image markers as guides to set a protocol with the limiting boundary criterion for brain regions depicting on each

image slice and thereby identifying all the true multiple connected brain regions resulting in more sophisticated and accurate segmentation results.

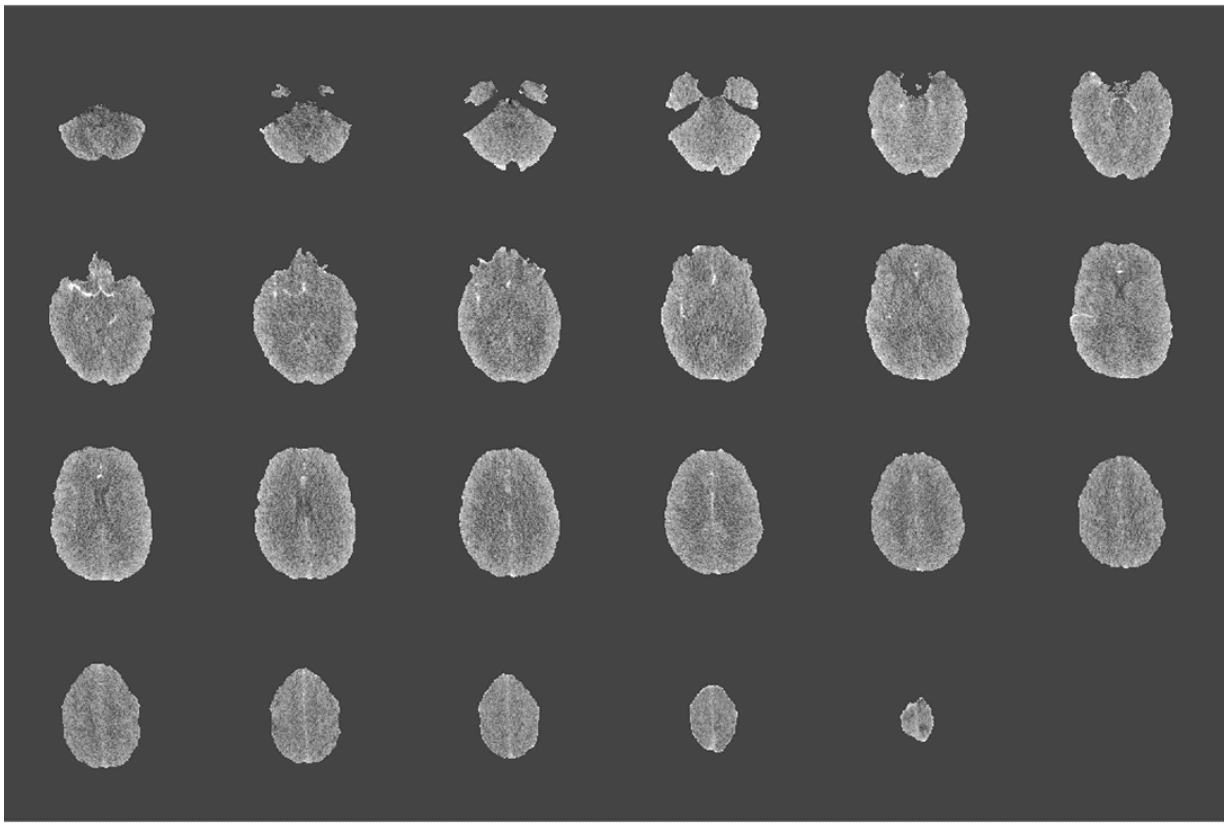
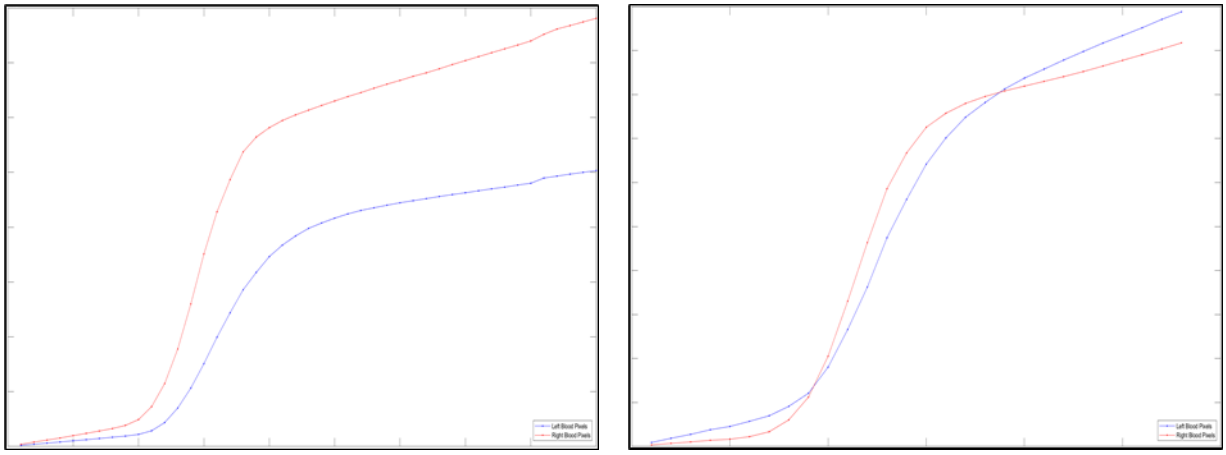


Figure 4-7: Illustration of proposed segmentation scheme using image markers and consecutive mapping technique for a sample brain series.

Figure 4-8 illustrates and compares the difference of the cumulative blood volume between the left and right hemispheres in two cases. In case (a), there is a clearly big difference in the cumulative blood flow and volume between two hemispheres throughout the CTP image acquisition period, which indicates the presence of major LVO. Thus, applying EMT to remove blood clot and resume blood supply can help balance blood in both hemispheres. Since this patient can benefit from the EMT and thus receive good clinical outcome (in class-0 of mRS). However, case (b) has relatively small difference of cumulative blood flow or transit time in both hemispheres. Thus, the underlying reason for the poor prognosis is not primarily caused by LVO

or unbalanced blood supply. The clinical result shows that this patient does not benefit from EMT and is classified into class-1 of mRS in this study.



(a) (b)

Figure 4-8: Comparison between two cumulative blood flow curves in left and right hemispheres of the brain, where case (a) is classified to 'class-0' and case (b) is classified to 'class-1' of mRS.

By analyzing all 31 testing cases in our dataset, Table 4-2 shows and compares the number of input features that are generated using PCA algorithm and used to train ML models, as well as the classification performance (AUC values) of 5 models. The corresponding ROC curves are presented in Figure 4-9. The results show that Model-III built using the features computed from the subtracted blood flow curves (related to the transit time for contrast agent wash-in and wash-out) between two hemispheres of the brain produces the highest AUC values as compared to Model-I and Model-II. Using KNN and SVM ML methods, Model-III.2 (KNN) and Model-III.1 (SVM) yield AUC=0.878±0.077 and 0.846±0.078, respectively.

Table 4-2: Summary of the number of PCA features used in various ML models and their corresponding classification performance in terms of AUC.

Model	Number of Features	AUC
Model-I	1	0.772±0.084
Model-II.1: SVM	5	0.746±0.089
Model-II.2: KNN	5	0.607±0.103
Model-III.1: SVM	4	0.846±0.078
Model-III.2: KNN	4	0.878±0.077

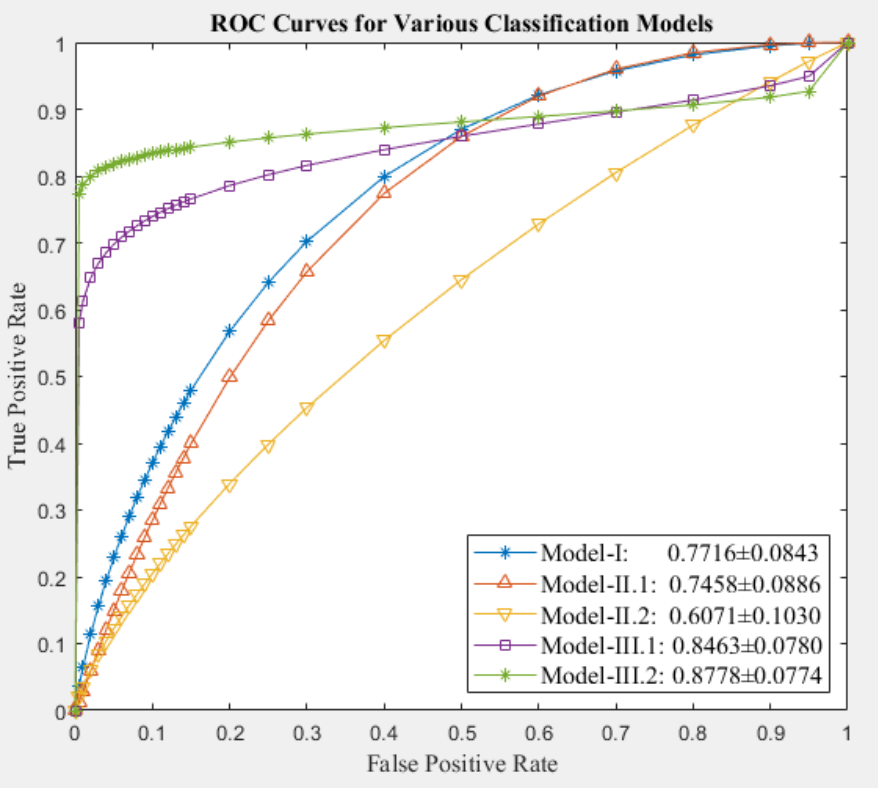


Figure 4-9: Comparison of various ROC curves generated using 5 models to classify between two mRS classes.

Table 4-3 illustrates 5 confusion matrices of 5 models. Based on these matrices, a set of the computed performance indices including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), are summarized in Table 4-4. From these

tables, one can observe that like AUC value assessment (Table 4-2), the highest classification accuracy is obtained for Model-III in which Model-III.2 (KNN) achieves the highest prediction accuracy of 90.3%.

Table 4-3: Summary of confusion matrices of various ML models to classify between mRS classes.

Model	Actual	Predicted	
		Positive	**Negative**
Model-I:	Positive	11	5
	Negative	4	11
Model-II.1: SVM	Positive	13	3
	Negative	6	9
Model-II.2: KNN	Positive	10	6
	Negative	5	10
Model-III.1: SVM	Positive	13	3
	Negative	4	11
Model-III.2: KNN	Positive	16	0
	Negative	3	12

Table 4-4: Summary of several performance indices for various ML models.

Model	Accuracy (%)	Sensitivity	Specificity	PPV	NPV
Model-I:	70.97	0.69	0.73	0.73	0.69
Model-II.1: SVM	70.97	0.81	0.60	0.68	0.75
Model-II.2: KNN	64.52	0.63	0.67	0.67	0.63
Model-III.1: SVM	77.42	0.81	0.73	0.76	0.79
Model-III.2: KNN	90.32	1.00	0.80	0.84	1.00

4.4 Discussion

In this study, we investigate the feasibility of developing and applying new quantitative image markers or ML models to predict prognosis of AIS patients at an early stage using pre-intervention brain CTP images. The study has several unique characteristics and contributions. First, we apply several novel image processing algorithms to develop a new CAD scheme that can be applied to real clinical images with varying imaging scanning conditions. Based on our literature search, no similar CAD schemes are available to date. Our CAD scheme can automatically perform following tasks including (1) organizing and matching all CTP image slices in a correct order of scanning sequences, (2) segmenting brain volume and contrast-enhanced blood volume in all CTP image slices, (3) generating two cumulative blood flow curve diagrams of left and right hemispheres of the brain, and (4) computing image features related to the blood transit time and velocity in 3 normalized phases of CTP scanning sequences. To increase reliability and user confidence to the CAD scheme, a unique interactive GUI is designed and used (as shown in Figure 4-3). As a result, any possible image processing errors (i.e., brain or blood volume segmentation errors) can be visually observed and corrected either automatically by performing the pre-installed correction functions in the GUI or manually by the user's hand drawing or new boundary condition setting. Although we did not test the GUI functions in this study, the similar GUI tool developed in our previous study [103] has been tested and used by the clinical researchers in the Department of Neurology of OUHSC to provide quantified percentage of blood leakage volume in aneurysmal subarachnoid hemorrhage (aSAH) patients using brain CT images, and then predict clinical outcome of the patients as reported in our previous publications [108], [52].

Second, due to the potential presence of unilateral blood clot (LVO) in AIS patients, the transit time and velocity of the blood contrast flow rate (wash-in and wash-out) may vary between two

hemispheres of brain. Thus, asymmetrical blood flow rate or pattern in two hemispheres of the brain provides a potentially useful image marker to predict AIS prognosis. Instead of subjective assessment of asymmetrical blood flow rate or transit time by the neuro-radiologists, which is qualitative and has large inter-reader variability, our CAD scheme computes several quantitative features to assess asymmetrical blood flow rate and patterns. Our data analysis results demonstrate that using the absolute cumulative disparity value has higher discriminatory power to predict prognosis of AIS patients with AUC=0.772±0.084 (Table 4-2), which is significantly higher than random guess (AUC=0.5). In addition, our GUI tool (Figure 4-3) provides clinicians (i.e., neuro-radiologists) a visual-aid tool to examine or monitor transit time or velocity of blood flow and the final difference of cumulative amount of blood flow in the left and right hemispheres of the brain.

Third, another advantage of developing CAD scheme is to compute multiple features. Then, ML methods can be applied to select optimal features and build multi-feature fusion models aiming to achieve the increased prediction performance than using a single optimal feature or marker. In this study, we investigate two sets of features. As shown in Figure 4-4, one includes cumulative blood flow or contrast agent transit time slopes related features computed separately from the left and right hemispheres of the brain, while another one includes features computed from the subtracted cumulative blood flow curve of two hemispheres. The study results (Table 4-2) demonstrate that multi-feature fusion models yield significantly higher prediction performance than using a single image feature or marker ($p<0.01$). Model-III also yields significantly higher performance than model-II for using both SVM and KNN learning methods. It indicates that using the absolute difference curve of blood flow between two hemispheres carriers more discriminatory information or power to train ML models than using the features computed separately from two hemispheres. It also reduces the number of features (Table 4-2), which can help improve

robustness of ML models. Additionally, performing paired t-test showed that 4 out of 6 analyses have a significant statistical performance ($p < 0.05$) amongst the multi-feature fusion models. For instance, the best performing Model-III.2 (in terms of accuracy and AUC), has statistically significant performance as compared to Model_II.1 and Model-III.1. More details related to the p-values computed between the various ML methods are summarized in Table 4-5.

Table 4-5:Comparison of p-value between various ML models.

P-value	Model-I	Model-II.1	Model-II.2	Model-III.1	Model-III.2
Model-I		0.238689	0.20936	0.218763	0.234386
Model-II.1			0.009978	0.382433	0.012067
Model-II.2				2E-05	0.25409
Model-III.1					2.65E-07
Model-III.2					

Fourth, many radiomics studies have recently reported that radiographic images (i.e., CT) contain useful phenotype image features, which highly associate with prognosis of cancer patients [97], [98]. In this study, we demonstrate the feasibility of identifying and applying radiographic image features or markers computed from brain CTP images to potentially phenotype AIS patients and predict their prognosis. This study supports and expands radiomics concept to more broad clinical application fields. The success of our approach to develop new quantitative image markers or prediction models will eventually provide clinicians a new decision-making supporting tool to assist them more accurately stratifying AIS patients for choosing and applying optimal treatment methods (i.e., EMT) at an early stage aiming to reduce patients' mortality and morbidity rates.

Despite the encouraging results, we also recognize that this study has several limitations. First, only a small dataset is used in this study. Distribution of cases belonging to each mRS class (0 to

6) is also not uniform, which restricts us to divide the dataset into only two classes to represent good and poor prognosis. Due to the small dataset, the computed AUC values have relatively big standard deviation (Table 4-2) indicating the relatively lower confidence level. Additionally, small dataset may not well-represent the general AIS population in the real clinical environment. Thus, the performance and robustness of our CAD scheme and image marker or ML models need to be further tested using larger and independent study cohorts in future studies. Second, the simple case normalization to compensate the different numbers of the CTP scanning sequences and division of cumulative blood volumes into three equal phases may not be optimal. A more dynamic or adaptive division of cumulative blood volume phases should be investigated and compared in future studies. Third, to fully use radiomics concept, more image features need to be explored and computed to identify more discriminatory information to improve performance and robustness of ML models. Last, this study uses a simple threshold and labeling algorithm to segment brain and blood volume. Currently, deep learning technology has been applied to segment the targeted regions of interest in medical images [109], [110]. Using deep learning method may help achieve higher accuracy in segmentation of brain and blood volume using CTP images in future.

In conclusion, this is a preliminary and proof-of-concept study to develop new quantitative image markers to classify AIS patients based on mRS severity using a set of bilateral asymmetrical image features computed of CTP images between the left and right hemispheres of the brain. The study demonstrates the promising results when applying the CAD scheme and ML model to a set of diverse clinical cases with different mRS distribution and varying CTP imaging scanning protocols. Based on the foundation built in this study, new research effort can follow to further validate these quantitative image markers and conduct prospective clinical studies in the future.

5 Applying Quantitative Radiographic Image Markers To Predict Clinical Complications After Aneurysmal Subarachnoid Hemorrhage: A Pilot Study

5.1 Introduction

Aneurysmal subarachnoid hemorrhage (aSAH) is a medical emergency associated with high mortality, morbidity, and significant healthcare burden[111]. Nearly 30,000 people are affected by aSAH every year in the United States[112]. Recent studies suggest early brain injury (EBI) and delayed cerebral ischemia (DCI) that results from several pathophysiological processes are major determinants of mortality and morbidity associated with aSAH patients. EBI typically occurs within 72 hours of aSAH while DCI is commonly seen 4-14 days of aSAH onset. Thus, early and quick diagnosis and prediction of aSAH prognosis is important to more effectively treat aSAH patients in clinical practice.

EBI occur secondary to one or more of the following: acute hydrocephalus (HCP), transient cerebral oligemia, and systemic inflammatory response. EBI causes significant morbidity and mortality[49]. Despite this known associative role of EBI in determining clinical outcome, absence of objective radiological image biomarkers limits its role in assessing disease severity and predicting prognosis. The commonly used modified Fisher scale (mFS) identifies amount of blood in subarachnoid space and ventricles, although easy to use and able to predict cerebral vasospasm and DCI, it is subjective with high inter-rater variability[50],[51]. Quantifying blood using above-mentioned parameters have decent ability to predict occurrence of DCI and long-term clinical outcome, but it is semi-automated and also operator dependent[52]. In addition, recent studies reported DCI as a complex process, not fully explained by intracranial blood or HCP[113]. Hence,

researchers have investigated if EBI could be radiologically assessed through global cerebral edema (GCE) that might incorporate some of the above-mentioned complex process to more objectively predict clinical outcome. Recently, a semi-quantitative non-automated scoring system has been investigated to assess GCE using Subarachnoid Hemorrhage Early Brain Edema Score (SEBES) that is based on the visibility of sulci at two pre-determined levels, which may predict DCI with high accuracy[55]. Additionally, quantification of cerebral edema using semi-automated or automated segmentation of sulci volumes has shown to be an important marker for analyzing EBI after aSAH[53],[54].

As described above, in order to detect or quantify severity of aSAH and predict its prognosis or clinical outcomes, non-contrast computed tomography (CT) imaging is typically applied to scan patients' brain due to high efficiency of CT imaging scan to detect leaked blood on brain tissue. However, subjectivity of neuroradiologists in reading brain computed tomography (CT) images plays a significant role in assessing EBI radiologically and hence prone to predictive errors. It is likely that many useful imaging features depicting on brain CT images acquired at admission of patients and subsequent CT scans during hospitalization may or may not be correctly identified through visual inspection and subjective interpretation of neuroradiologists alone. Applying an automated program to detect and segment brain CT images has potential to provide a standardized and unbiased estimation of targeted radiological image parameters. Hence, in this exploratory study we develop and test a novel and automated radiographic imaging software to accurately quantify image features that pathophysiologically contribute to short-term (in-hospital) and long-term clinical outcome. Successful development of such radiological image biomarker will help in the early prediction of possible clinical outcome of aSAH patients. The hypothesis used to develop current image analysis software is based on the two studies[55], [53] that used automated program

to segment and quantify a pre-determined level above lateral ventricle to perform a volumetric assessment of sulci, white matter (WM), gray matter (WM), and extra-parenchymal blood (EPB). Hence, the objective of this study was to develop and test a computer-aided detection (CAD) scheme to automatically segment brain regions and generate radiographic imaging biomarkers computed from brain CT images to predict various clinical outcomes after aSAH.

5.2 Materials and Methods

5.2.1 Study Population:

This study was approved by Institutional Review Board at the University of Oklahoma Health Sciences Center (OUHSC). All patients used as study subjects had previously signed written informed consent forms for their participation in the long-term follow-up portion of the study during their 1-year clinical follow-up as discussed in our previous studies[52], [114]. In brief, CT scans of 59 patients admitted with the diagnosis of aSAH at the OU Medical Center between the years 2014 and 2016 were retrospectively analyzed for this study. The inclusion criteria for the study subjects required patients diagnosed with aSAH and a CT scan acquired within 48 hours of symptom onset. Multiple CT scans of the head were acquired to different clinical indications during the hospital course. However, for this study, we only used two non-contrast CT scans acquired during patients' admission and day 10-14 after hospitalization, which are termed as the first and last CT scans in the rest of this paper. The 10-14 days were chosen as it is generally accepted to be the end of DCI period (prior-to-discharge).

Following 4 short-term clinical outcomes during hospitalization were retrospectively studied: (i) Clinical vasospasm (CVSM) – defined as patients developing clinical symptoms attributable to cerebral vasospasm after exclusion of other confounding causes, (ii) DCI – defined as occurrence of CVSM and/or radiological evidence of cerebral infarction, (iii) HCP – assessed on admission

CT scan using age-corrected bicaudate index[115], and (iv) ventriculoperitoneal shunt (VPS) – placement during or after discharge as clinically determined. Additionally, clinical outcome assessed during post-discharge clinic visit includes two indices namely, (1) a modified Rankin score (mRS) that assesses physical disability ranging 0 to 6 with higher values indicating worse outcome and (2) a Montreal cognitive assessment (MoCA) index that assesses cognitive disability which is scored 0-30 with lower score indicate presence of cognitive dysfunction. These generate 4 long-term clinical outcomes defined in this study.

Due to the small dataset of 59 patients, we divided each clinical outcome into two categories of good or poor outcome, as discussed below. The clinical measures DCI, CVSM, HCP, and VPS, were categorized into two classes of either presence (poor) or absent (good). The mRS is an ordinal disability score categorized into seven scales (from 0 = no symptoms to 6 = dead). We divided mRS into two categories: 0 to 2, independent functioning and 3 to 6 as dependent outcome or death. A MoCA score of 26 or greater is considered normal. Thus, we divided MoCA into two categories a MoCA score of 26 or over (normal cognition) and a MoCA score of less than 26 (presence of cognitive disability). The mRS and MoCA were performed in patients at ~3-month (3M) and ~1 year (1FU) follow-up visit. The distribution of the patients for the above eight measures along the demographics was summarized in Table 5-1.

Table 5-1: Demographics and Clinical Characteristics of patients (n = 59). mRS: class-1: 0-2; class-2: 3-6; MOCA: class-1: ≥ 26; class-2 < 26; SD – Standard Deviation; IQR – Interquartile range, 3M – 3 months, 1FU – 1 year follow-up.

Clinical Characteristics			Number of cases (n)	
			First scan	Last scan
Sex			Male: 20; Female:39	-
Age: (Min 20; Max 83)				
Mean ± SD	Median (IQR)		52.7 ± 12.9 53 (44, 60)	-
DCI (Yes/No)			59 (33/26)	57 (33/24)
CVSM (Yes/No)			59 (23/36)	57 (23/34)
HCP (Yes/No)			39 (15/24)	37 (15/22)
VPS (Yes/No)			53 (13/40)	51 (13/38)
mRS at 3M (class-1/class-2)			41 (27/14)	40 (26/14)
MOCA at 3M (class-1/class-2)			41 (19/22)	40 (19/21)
mRS at 1FU (class-1/class-2)			36 (22/14)	34 (20/14)
MOCA at 1FU (class-1/class-2)			36 (16/20)	34 (14/20)

5.2.2 Flow Diagram of the Proposed CAD Scheme:

CAD scheme is divided into two phases namely, image processing phase and data analysis phases. Image processing phase is common, while data analyses phase changes depending on the clinical measure being evaluated. During image processing phase, each three-dimensional (3D) input brain CT scan is processed to (i) automatically segment volumetric region of the intracranial brain, (ii) label each voxel to different subcategory, and (iii) compute corresponding radiographic image features. While in data analysis phase, the computed image features associated with a specific type of scan (first or last) coming from image processing phase are combined with class

labels of each clinical measure to perform classification analysis. In this study, we assessed classification performance of our CAD scheme for each of eight clinical measures as shown in table 1 independently using both first and last CT scans. As a result, 16 different classification analyses were performed belonging to 8 clinical measures for both the first and last CT scans. A sample representation of the detailed flow diagram of the proposed CAD scheme highlighting individual steps in each phase is shown in Figure 5-1. A more detailed explanation of each step of CAD scheme are discussed in the subsequent sections.

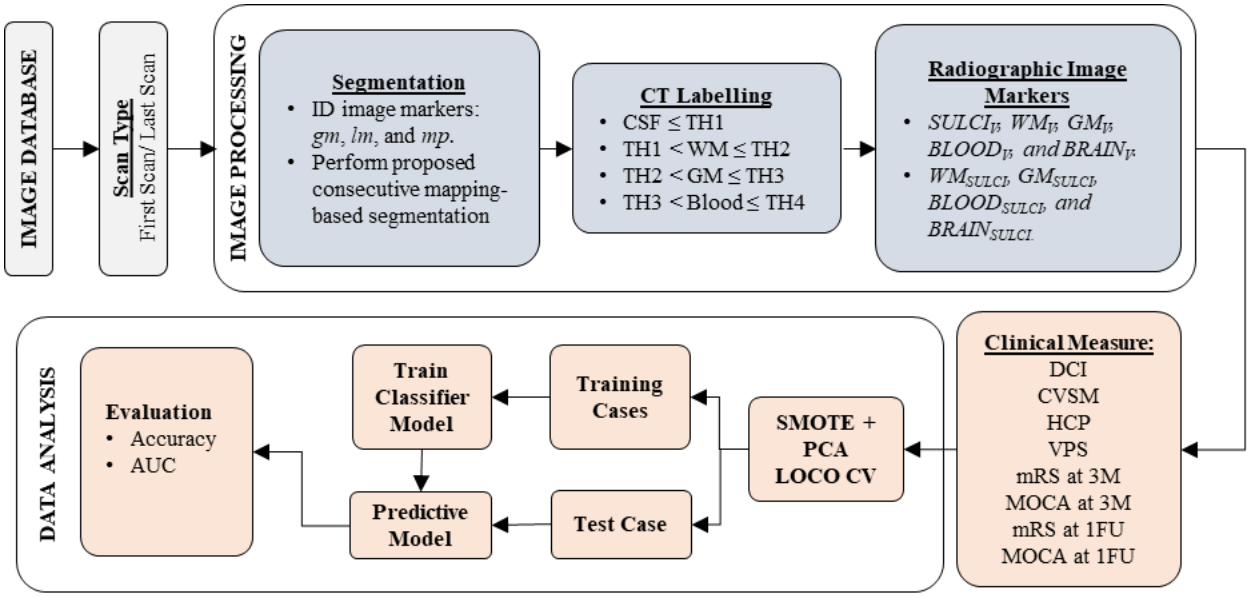


Figure 5-1: A detailed flow diagram of each step of the proposed CAD scheme categorized by the image processing and the data analysis phases.

5.2.3 3D Brain Segmentation and CT Labeling:

A two-stage consecutive mapping-based algorithm is implemented in this CAD scheme to achieve optimal segmentation results[116]. During the first stage, for each 3D CT scan sequence, CAD scheme applies an initial image pre-processing step to identify three image markers namely, global minimum (gm), local minimum (lm), and maximum peak (mp) of intracranial brain region. These image markers will represent slice indices of the current CT sequence to act as reference points to guide CAD segmentation algorithm. To achieve this, CAD performs Otsu thresholding

to filter the largest single-connected intracranial brain region belonging to each of the 2D-image slices. Next, based on a comparison of segmented brain volumes, the slice with the largest volume is considered mp, the slice with the smallest volume is regarded as gm, and the slice with the smallest volume in the opposite direction concerning the mp slice gm slice is marked as lm. The initial segmentation results at the end of the first stage are still prone to errors regarding leakages and identification of multiple brain regions per slice. Thus, in the second stage, a fine-tuning technique in reference to the neighboring slices is applied to obtain precise segmentation. Correction in this stage begins with the mp slice as the intracranial brain region in this slice is the most significant, single connected component and wholly enclosed within the skull, facilitating accurate segmentation. Then, the fine-tuning procedure scans either in the left or right direction, applying the consecutive mapping technique. During this mapping-based correction, the segmentation results from the prior slice are used to define the limiting boundary criterion to avoid segmentation leakages and identify multiple connected brain regions. The limiting boundary criterion performs morphological dilation on prior segmentation results to define current slice boundary limits, which prevents segmentation leakages. In the case of multiple connected regions, each individual region is examined with prior segmentation results and constrained to enclose within the limiting boundary for inclusion in the current slice segmentation. To summarize the second stage of the segmentation method, the process begins and considers the segmentation at mp slice as the reference, then continues using this consecutive mapping-based algorithm until either a gm slice or lm slice is reached in both directions covering all the image slices.

After successful brain region segmentation, based on the standardized calibration of Hounsfield Units (HU) of CT images, each CT voxel inside the brain is subcategorized into four regions, namely, Cerebrospinal fluid (CSF), White Matter (WM), Grey Matter (GM), and blood

using multi-thresholds of Hounsfield Units (HU). In this study, the default threshold (TH) values defined for subcategorization of voxels were TH1: 20, TH2: 32, and TH3: 50. Any voxel with an HU value less than or equal to TH1 is identified as CSF during the CT labeling protocol. In contrast, all the voxels with HU values greater than TH3 are labeled as blood. Then, the voxels with HU values greater than TH1 and less than or equal to TH2 are labeled as WM. Finally, the voxels with HU values greater than TH2 and less than or equal to TH3 are labeled as GM. To summarize, CT labeling protocol: (1) CSF ≤ TH1, (2) TH1 < WM ≤ TH2, (3) TH2 < GM ≤ TH3, (4) TH3 < blood was applied to categorize each voxel into one of the four regions. Following CT Labelling, volumetric analysis of these four regions was performed to identify and compute radiographic image features. The analysis was conducted only on image slices between the level of lateral ventricles until the top of the skull. This pre-determined level was picked from the previous study[55], where the presence of sulci is analyzed using a qualitative assessment for association with aSAH. All the CSF regions automatically identified by the CAD scheme are considered as sulcal regions as CSF is filled within the sulcal space.

5.2.4 Computing Radiographic Image Features:

CAD scheme is applied to compute the image features from CT images slice-by-slice. The three-dimensional (3D) volumetric representation of each image feature is then computed by weighted summation of feature values computed in all corresponding CT image slices. For each CT label (i.e., Sulci, WM, GM, Blood, and Brain), an initial slice-based representation of volumes is computed using the number of the segmented voxels (N_{Label}) and slices parameters such as voxel length (V_L), voxel breadth (V_B), and slice thickness (S_T) as shown in equation 5.1.

$$Label_V|_{slice} = N_{LABEL} \times V_L \times V_B \times S_T|_{slice} \quad (5.1)$$

Next, a summation of these slice volumes was performed to yield 3D volumes of each labeled category as shown in equation 5.2.

$$Label_V = \sum_{slice_{k=m}}^{n} Label_V|_{slice_k} \tag{5.2}$$

Above process is repeated independently for each of the CT labels to obtain corresponding volumetric features. For instance, to compute the 3D volume of the labeled sulci, CAD scheme first computes the segmented sulcal volumes depicting on each slice separately and then combine the results of all the slices to obtain the 3D volume of sulci as shown in equations 5.3 and 5.4.

$$Sulci_V|_{slice} = N_{Sulci} \times V_L \times V_B \times S_T|_{slice} \tag{5.3}$$

$$Sulci_V = \sum_{slice_{k=m}}^{n} Sulci_V|_{slice} \tag{5.4}$$

In this way, the first group of features includes volumes of 5 labeled categories namely, sulci, white matter (WM), gray matter (GM), blood, and brain tissue, are computed, which are represented using a "V" as subscript following each label as follows: (i) $SULCI_V$, (ii) WM_V, (iii) GM_V, (iv) $BLOOD_V$, and (v) $BRAIN_V$.

Additionally, CAD scheme also computes the second group of features that includes 4 features obtained by calculating the proportion $SULCI_V$ with respect to the volumes of remaining labels (WM_V, GM_V, $BLOOD_V$, and $BRAIN_V$) and represented with a "*SULCI*" as a subscript following each label as follows: (i) WM_{SULCI}, (ii) GM_{SULCI}, (iii) $BLOOD_{SULCI}$, and (iv) $BRAIN_{SULCI}$, respectively. For instance, the proportion of sulci with respect to WM is represented using WM_{SULCI} and is computed as shown in equation 5.5.

$$WM_{SULCI} = \frac{Sulci_V}{WM_V} \tag{5.5}$$

As a result, a total of nine radiographic image features, including five features from the first group and four features from the second group, are extracted and computed from 3D CT slices of each case.

5.2.5 Building Machine Learning Model and Evaluation:

We then build and test several multi-feature fusion-based machine learning (ML) models to classify prognosis of the study cases into two classes. Specifically, we investigated eight clinical measures (as shown in table 1) for two CT scans (first and last), resulting in a total of 16 different independent analyses of case classification. Due to the wide range of data analysis, we examined the feasibility of applying three popular supervised ML architectures namely, support vector machine (SVM), k-nearest neighborhood (kNN), and logistic regression for the classification tasks in this study. Based on our study experience, a polynomial kernel is used in the SVM model and k = 10 (neighbors) is applied in the kNN model.

In order to build robust and highly performed ML models, following three critical issues or challenges need to be considered and addressed, which are: (i) imbalance of cases between two classes, (ii) a scope for redundancy in the initial feature pool, and (iii) a relatively small dataset. First, there is a noticeable imbalance between the number of samples belonging to each class regarding each clinical measure in our dataset. For example, in the case of CVSM, we have a total of 59 samples with the first CT scan, and the number of samples belonging to class-1 vs. class-2 is 23(39%) vs. 36(61%), respectively. More details related to the class imbalance situation regarding each of the 16 analyses can be observed in table 1. To address this issue and minimize its impact on training ML model, we applied a Synthetic Minority Oversampling Technique (SMOTE) algorithm[117] to rebalance the number of samples in each class. The vital point of

applying SMOTE is that it introduces synthetic data samples by interpolation between some minority class instances within a specified neighborhood. The fundamental purpose of the SMOTE is to generate synthetic data samples belonging to the minority class label using the interpolation technique of minority class instances within a specified neighborhood. For instance, if we need a total of *N* new synthetic samples to balance the minority class, according to a distance metric, we first select *K* samples belonging to the minority class from the training set. Then, among these *K* instances, *N* instances are selected randomly for computing the new instances by interpolation. A sample illustration of SMOTE can be seen in Figure 5-2 in which *u* represents the minority class label, and [*u1, u2, u3, u4]* represents the *N* instances selected. In contrast, the points [*v1, v2, v3, v4]* represent the synthetic data generated using interpolation. Second, to potentially avoid overtraining and reduce the redundancy of the input features, we applied principal component analysis (PCA). The PCA is set to generate a new principal component feature vector with a variance rate of 95%. Third, a leave-one-case-out (LOCO) based cross-validation method is adopted to train and evaluate each ML model due to the small dataset. A LOCO validation method maximizes learning power, while minimizing the case partition and testing bias as demonstrated in previous studies (i.e.,[87], [118]).

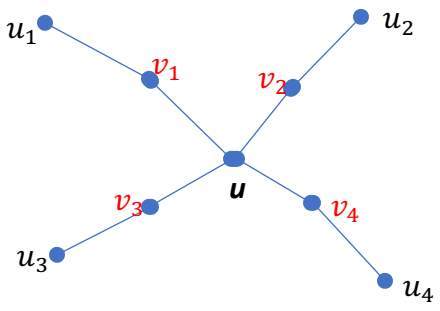


Figure 5-2: A sample illustration of generating synthetic data of the minority class label using the interpolation of SMOTE algorithm.

We embedded both SMOTE and PCA algorithms into the LOCO cross-validation process during training and testing each ML model. Specifically, in each LOCO training and testing iteration, we first take out one case from the training set as independent testing case. Second, SMOTE algorithm is applied to generate synthetic data to balance the minimum class instances. Third, PCA is applied to the training samples with a variance rate of 95%, thereby, generating a new feature vector of reduced length. Fourth, the ML model is trained using all training samples (including the synthetic samples). Last, after the ML model is trained, the model is applied to the testing case to generate a classification score. This SMOTE and PCA embedded LOCO training and testing iteration process is repeated 59 times for testing all cases in our dataset. As a result, 59 classification scores generated by each ML model are recorded.

To evaluate performance of each ML model, we used following two steps. First, a receiver operating characteristic curve (ROC) is constructed from the classification scores using a maximum likelihood-based ROC curve fitting program (ROCKIT, http://www-radiology.uchicago.edu/krl/, University of Chicago), and the area under the ROC curve (AUC) is computed and used as an index to evaluate and compare the performance of each ML model to classify between two classes. Second, we apply an operating threshold on the classification scores (T = 0.5) to divide all testing cases into two classes (score ≤ 0.5: 'Class-1'; score > 0.5: 'Class-2'). From the classification results, several confusion matrices corresponding to different ML models are generated, which are used to compute various performance indices of class classification (i.e., classification accuracy (ACC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV)).

5.3 Results

Figure 5-3 shows one sample of case segmentation results applying the proposed consecutive mapping-based algorithm in our CAD scheme. The sequence of images is organized from the top left to the bottom right in a left to right fashion. CAD scheme identifies 28 CT image slices depicting visible brain regions with varying size in the first 25 CT slices. In this case, CAD identifies slice-11 as the mp slice with the most significant brain area. In contrast, slice-1 and slice-25 are assigned as gm and lm, respectively. During the fine-tuning stage from mp towards the gm (left direction), we can notice that the guided limited boundary criteria defined from the prior slice segmentation have assisted in both avoiding leakages and identifying multiple connected brain regions (noticed in slices 1 through 7). Additionally, during fine-tuning stage from mp towards the lm (right direction), when the top of the skull is reached at slice-26, segmentation is not performed for slices 26 through 28 as it is not required.

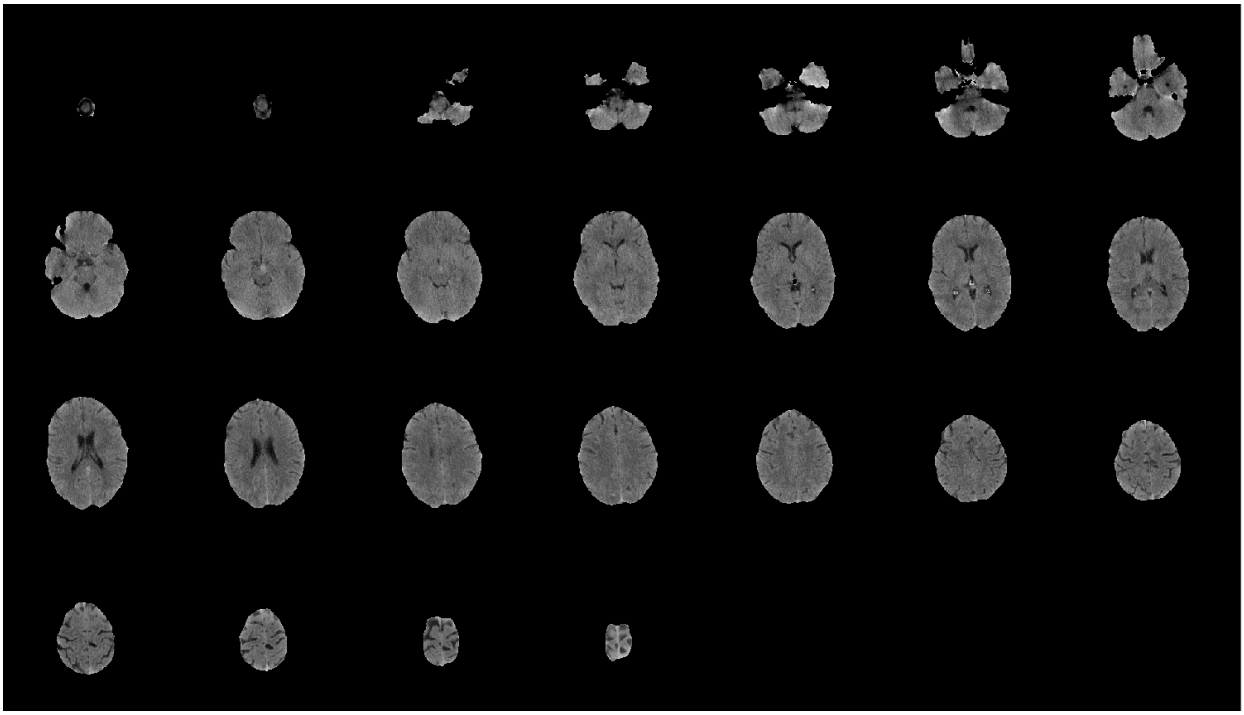


Figure 5-3: A sample case segmentation results using the proposed consecutive mapping-based algorithm. The sequence of images is organized from the top left to the bottom right in a left to right fashion.

As explained earlier, in this study, image data analysis is conducted only on image slices between the level of lateral ventricles until the top of the skull. Thus, for each case, the last slice index depicting the ventricles is marked for reference. Then, all the images above the defined ventricle level until the top of the skull are selected for further analysis by CAD scheme. Figure 4 shows a sample series of images representing the CT labeled images within the defined region of analysis, including the reference slice consisting of ventricles. In Figure 5-4, slice-1 is only shown to illustrate the level of ventricles but not included in further analysis. Here, the images are labeled in greyscale ranging from dark to bright, in four levels where each level represents sulci, WM, GM, and Blood, respectively.

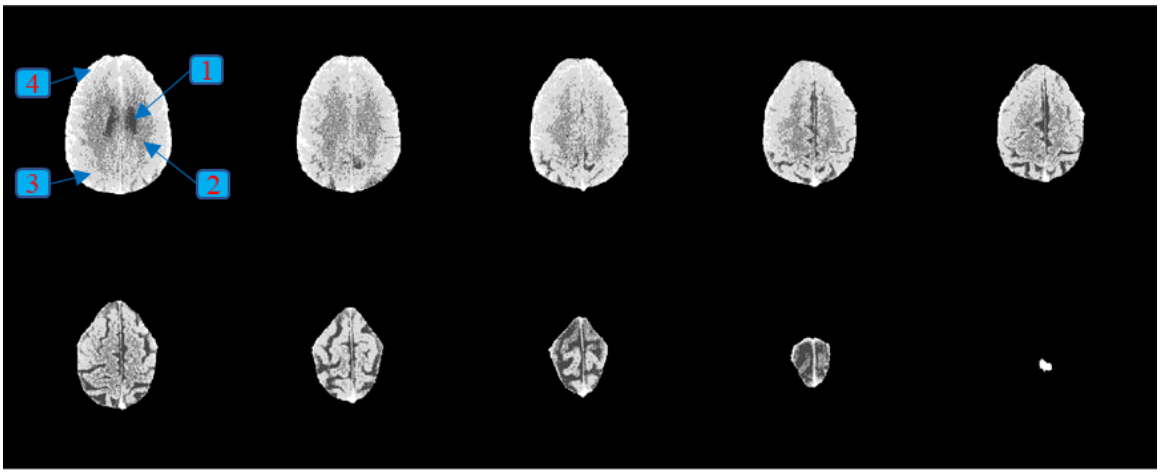


Figure 5-4: A sample case illustration of CT labeled images within the defined region of analysis. The first slice in the sequence is for ventricles reference purpose only and not included in the volumetric analysis of labeled regions. In the figure, greyscale ranging from dark to bright in four levels represents cerebrospinal fluid (1), white matter (2), gray matter (3) and blood (4), respectively.

Figure 5-5 and Figure 5-6 plot and compare the ROC curves to classify cases into two classes based on all eight clinical measures between the first and the last CT scan. The AUC values computed from the first CT scan range between 0.65±0.10 and 0.82±0.05, with the lowest value observed for classifying mRS at 1FU and the highest value for classifying DCI. Similarly, the AUC values for the last CT scan range between 0.62±0.07 and 0.86±0.07, with the lowest value observed for classifying DCI and the highest value observed for classifying both MOCA at 3M

and MOCA at 1FU. For simplicity, we separated the clinical measures into two groups, namely representing short-term and long-term clinical outcome. The short-term clinical measures are the complications that occur during hospitalization, including DCI, CVSM, HCP, and VPS. Whereas the long-term clinical measures are observed at three months and 1-year follow-up after hospital discharge, including mRS at 3M, MOCA at 3M, mRS at 1FU, and MOCA at 1FU. In terms of short-term measures, on comparison between the first and the last CT scan, we observe that majority of these complications (including DCI, CVSM, HCP) are more accurately predicted using the first CT scan. In contrast, only VPS can be observed to have a higher accuracy using the last CT scan. On the contrary, to predict long-term measures (mRS and MOCA at 3M and 1FU), classification accuracy is higher using the last CT scan than using the first CT scan, which indicates impact of treatment during hospitalization of the patients. More details regarding the best model selected for each of 16 data classification analyses and the performance metrics (AUC and ACC) are summarized in Table 5-2.

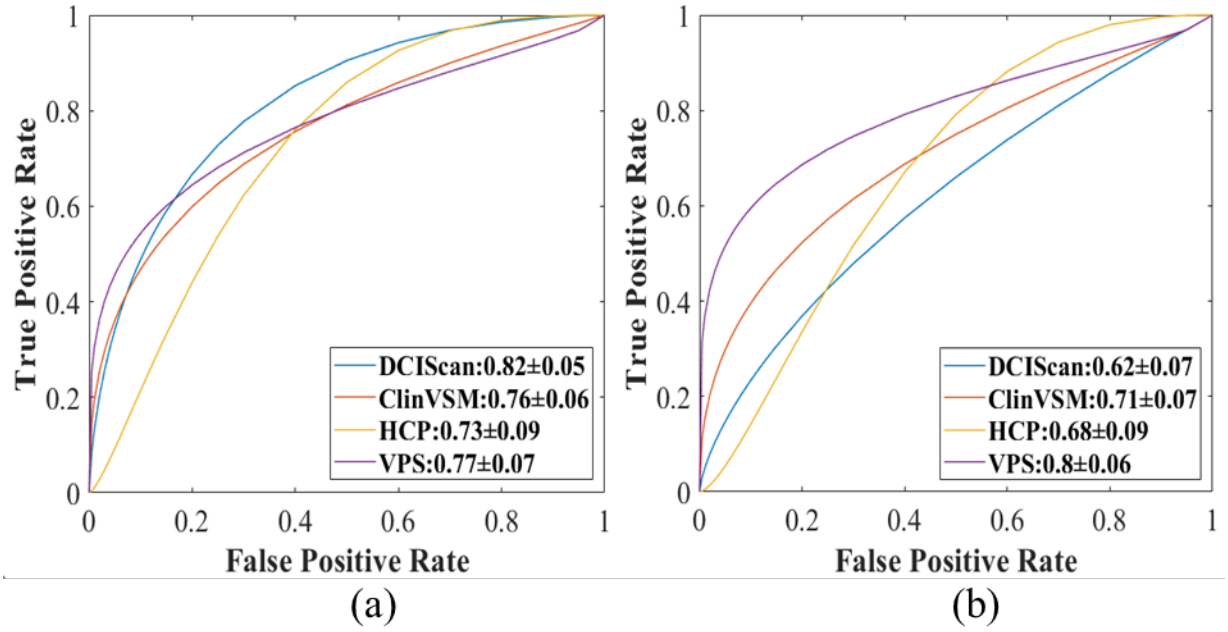


Figure 5-5: A comparison of ROC curves for the short-term clinical measures between using the first and the last CT scan. Where (a) represents the curves plotted from the first CT scan, (b) represents the curves plotted from the last CT scan.

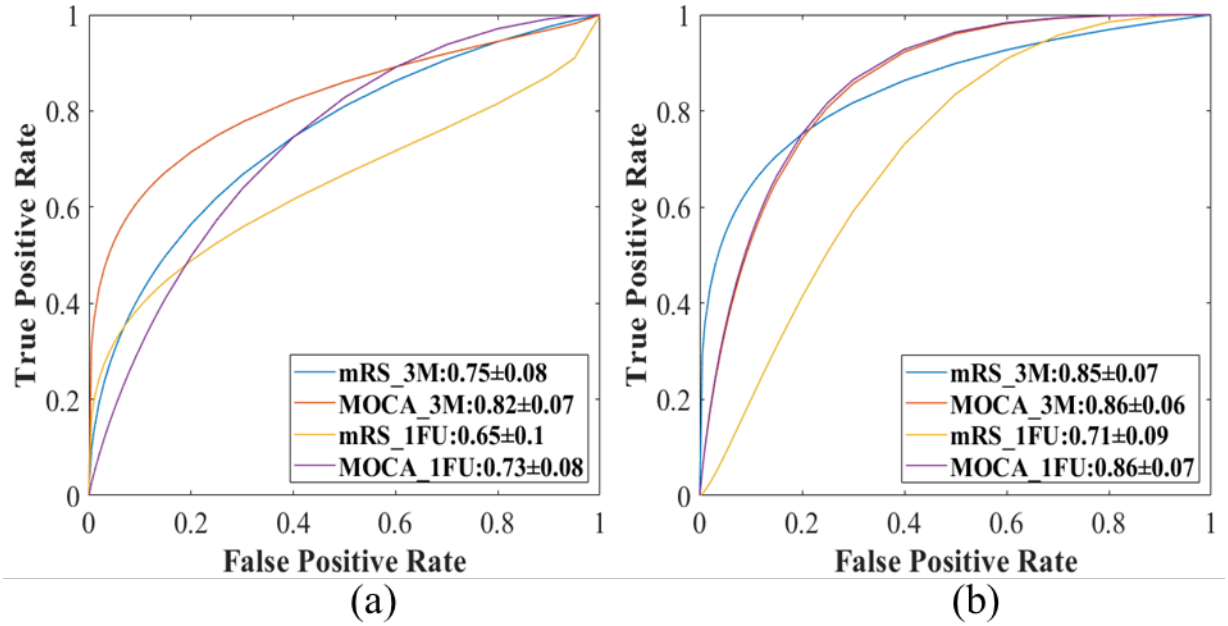


Figure 5-6: A comparison of ROC curves for the long-term clinical measures between the first and the last CT scan. Where (a) represents the curves plotted from the first CT scan, (b) represents the curves plotted from the last CT scan.

Table 5-2: A comparison of classification performance metrics and the best model selected for each of eight clinical measures between the first and the last CT scan.

Scan Type	Clinical Measure	Model	AUC±STD	ACC
First	DCI	SVM	**0.82±0.05**	**0.78**
	CVSM	SVM	**0.76±0.06**	**0.78**
	HCP	kNN	**0.73±0.09**	**0.72**
	VPS	SVM	**0.77±0.07**	**0.77**
	mRS at 3M	SVM	0.75±0.08	0.73
	MOCA at 3M	kNN	0.82±0.07	0.78
	mRS at 1FU	SVM	0.65±0.10	0.72
	MOCA at 1FU	SVM	0.73±0.08	0.64

Last	DCI	SVM	0.62±0.07	0.63
	CVSM	SVM	0.71±0.07	0.63
	HCP	SVM	0.68±0.09	0.68
	VPS	SVM	0.80±0.06	0.73
	mRS at 3M	SVM	**0.85±0.07**	**0.80**
	MOCA at 3M	Log Reg	**0.86±0.06**	**0.80**
	mRS at 1FU	SVM	**0.71±0.09**	**0.62**
	MOCA at 1FU	SVM	**0.86±0.07**	**0.79**

Additionally, a side-by-side comparison of both performance evaluation metrics including AUC and classification accuracy for each of the 16 data analyses in the study is shown in Figure 5-7. From accuracy values, we can notice that all the short-term measures are better performed using the first CT scan than the last CT scan. Whereas in the long-term measures, we observe that the majority of clinical measures or outcomes (including mRS at 3M, MOCA at 3M, and MOCA at 1FU) are better predicted using the last CT scan, and only mRS at 1FU is better predicted using the first CT scan. Last, based on 16 confusion matrices generated from the classification scores of their respective ML models, a set of the additional computed performance indices, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), are summarized and compared as shown in Table 5-3.

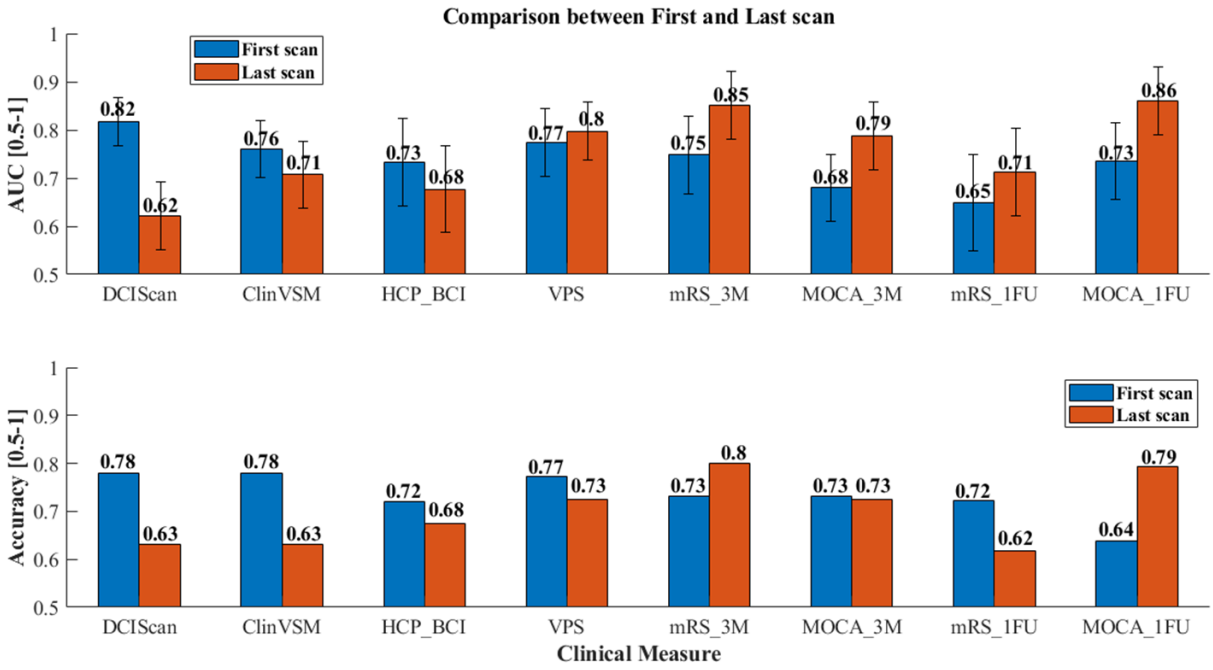


Figure 5-7: A side-by-side comparison of both the performance evaluation metrics AUC and accuracy for each of the 16 analyses.

Additionally, Table 5-4 shows two sets of AUC values along with the standard deviations of applying each of 9 single image features computed by our CAD scheme from the images of the first, and last CT scans to predict or classify mRS at 3-months among 41 aSAH patients. AUC values range from 0.52±0.10 to 0.73±0.09. The results indicate that many of these image features contain relatively higher levels of discriminatory power. For example, 4 image features yield AUC > 0.6 to predict mRS at 3-months using both CT images acquired at both the first and the last CT scans. However, the results also show that using an SVM model to optimally fuse multiple image features yields the highest classification performance (i.e., AUC = 0.85±0.07 using the images acquired from the last CT scan).

Table 5-3: A comparison of several other classification performance indices for each of eight clinical measures between the first and the last scan.

Scan Type	Clinical Measure	Sensitivity	Specificity	PPV	NPV
First	DCI	0.88	0.65	0.76	0.81
	CVSM	0.87	0.72	0.67	0.9
	HCP	0.73	0.71	0.61	0.81
	VPS	0.85	0.75	0.52	0.94
	mRS at 3M	0.96	0.29	0.72	0.80
	MOCA at 3M	0.74	0.82	0.78	0.78
	mRS at 1FU	0.91	0.43	0.71	0.75
	MOCA at 1FU	0.75	0.55	0.57	0.73
Last	DCI	0.79	0.42	0.65	0.59
	CVSM	0.83	0.5	0.53	0.81
	HCP	0.67	0.68	0.59	0.75
	VPS	0.77	0.71	0.48	0.9
	mRS at 3M	0.88	0.64	0.82	0.75
	MOCA at 3M	0.84	0.76	0.76	0.84
	mRS at 1FU	0.80	0.36	0.64	0.56
	MOCA at 1FU	0.86	0.75	0.71	0.88

Table 5-4:AUC values and standard deviations of applying each of 9 single image features computed by our CAD scheme to predict or classify mRS at 3-months using the images of the first and last CT scans.

Feature Number	First Scan	Last Scan
F1	0.66±0.10	0.57±0.09
F2	0.71±0.09	0.65±0.10
F3	**0.72±0.09**	**0.73±0.08**
F4	0.57±0.09	0.54±0.11
F5	0.64±0.09	0.58±0.10
F6	0.67±0.09	0.52±0.10
F7	0.56±0.10	0.72±0.08
F8	0.62±0.10	0.63±0.09
F9	0.62±0.10	0.66±0.09
Proposed Combined Model	**0.75±0.08**	**0.85±0.07**

5.4 DISCUSSION

This study demonstrates that sequential brain CT scans can play an important objective role to quantitatively predict both short-term and long-term clinical outcomes of aSAH patients. Although many clinical measures (i.e., 8 measures presented in Table 1) have been proposed and investigated, these measures are subjectively and qualitatively assessed by neuroradiologists in current clinical practice[119], which is a difficult task with potentially higher inter-rater variability. However, we propose developing this novel unbiased quantitative image analysis software, which can provide objective and quantitative radiological image marker or data analysis tool to predict

aSAH outcomes. Previous studies have demonstrated the advantages of using new image markers to facilitate clinical research and assist clinicians in their decision making process[120], [121]. In this study, we investigate and demonstrate the feasibility of developing and applying new quantitative radiographic image biomarkers using multi-feature fusion-based ML models to predict the prognosis of various clinical measures commonly used in current clinical practice after aSAH. The quantitative image markers enable to assess both the short-term and long-term clinical measures, which were predominantly observed between the first and last CT scans. To the best of our knowledge, no similar quantitative image markers have been developed and applied to predict variety of short-term and long-term clinical outcome measures of aSAH patients.

This study has several unique characteristics and contributions to develop and apply novel quantitative image markers. First, the image processing pipeline of the CAD scheme is fully automated and capable of performing various tasks including (i) accurate segmentation of intracranial brain volume using a novel mapping-based segmentation scheme, (ii) automated labeling to subcategorize each CT voxel into four clinically relevant brain regions, and (iii) computing relevant radiographic image features representing the volumetric characteristics of these defined brain sub-regions. Since all CAD steps mentioned above are performed automatically without any user intervention, the procedure is less time-consuming and has higher reproducible or robust results every time. Additionally, the CAD scheme is also visually interactive along with various functionalities that allow users to either monitor or adjust volumetric estimations. Even though none of these features were used in this study, they will be examined in future experiments by clinical professionals to improve the segmentation markings. These corrections will provide a ground truth of CT labeling, thereby, resulting in building more robust ML prediction or classification models.

Second, our study demonstrates the feasibility of identifying and applying radiographic image features or markers computed from brain CT images to potentially phenotype aSAH patients and predict the prognosis of various associated clinical measures. This study supports the radiomics concept and provides a new opportunity to expand the radiomics concept[122] to more broad clinical application fields. The study is influenced by an article[55] using qualitative markers evaluated by the clinicians as assessment in SAH patients and developed automatic quantitative image markers that are consistent and not prone to human error. Additionally, based on the results of this study, we have two observations namely, (i) the image markers extracted and computed from the admission (first) CT scan have potential to quickly predict the short-term clinical measures or outcome, which may help clinicians quickly decide optimal treatment options to improve clinical outcome, (ii) Whereas the image markers extracted from the 10–14-day (last) CT scan can more accurately predict the long-term clinical measures in aSAH patients. Such observation clearly supports the impact of patients' treatment on the long-term clinical outcome or recovery, which can assist clinicians developing optimal rehabilitation plan for the patients discharged from the hospital. As a result, our study results and observation indicate the potential advantages of developing new quantitative image markers, which has promising potential to provide clinicians (neuroradiologists) new decision-making supporting tools to assist them optimally treating initial brain injury and also mitigating secondary brain injury to reduce the morbidity and mortality of aSAH patients.

Third, unlike many previous studies to identify image markers based on single optimal image feature, the quantitative image markers developed and tested in this study are represented using ML model-generated classification scores. The ML models are built and integrated with multi-feature fusion methods that enable to capture the complementary image phenotype information.

This can achieve higher performance than using a single feature-based correlation method as demonstrated in many previous studies (i.e.,[87]). Additionally, many precautionary measures are taken while developing and training these ML models namely, (i) SMOTE algorithm is used to address and solve the class imbalance issue, (ii) PCA method is applied to generate optimal and non-redundant feature vector to achieve feature dimensionality reduction with a variance rate of 95%, and (iii) LOCO cross-validation method is used to maximize learning power of available dataset and eliminate bias in case partition. Both SMOTE and PCA are also embedded into LOCO cross-validation to further reduce training bias.

Fourth, although most previous studies used p-values or odds ratios to evaluate association between image or clinical markers and patients' prognosis or response to treatment, assessment results of association cannot be directly applied in the clinical practice to develop or establish precision medicine or personalized treatment paradigm. To overcome this limitation and increase clinical relevance, we use ROC-based data analysis method and build confusion matrices to evaluate performance or accuracy of the ML models to predict patients' prognosis or clinical measures. The reported AUC values along with classification sensitivity, specificity and positive or negative predictive values (PPV or NPV) have higher clinical significance to assist clinicians (neuroradiologists) making the optimal diagnosis and treatment decision applying to the individual patients in the clinical practice. Our study results with the highest AUC = 0.86±0.07 clearly demonstrate the feasibility or potential of developing and applying new quantitative image markers in future clinical practice.

Despite the encouraging results, we also recognize that this is only a pilot study with several limitations. First, the dataset assembled in this study is relatively small and may not sufficiently represent the varied spectrum of aSAH patient population. Thus, a more extensive and diversified

study cohort must be used to train/test a more robust CAD scheme. Second, simple thresholding is employed to perform CT labeling of various regions, which may not be ideal due to CT image noise. Although a more adaptive case-based thresholding may be investigated for better CT labeling. As the increase of dataset, an advanced deep learning model-based region segmentation method can also be developed and tested to segment brain CT images as demonstrated in one of our recent studies, which applies a modified deep leaning model to segment acute ischemic stroke (AIS) lesions using a publicly available AIS dataset. Third, we only compute and build a small feature pool with 9 features. As a dataset size increases, more radiomics types of features or deep learning model generated automated features can be computed and investigated. Fourth, even though the interactive graphical user interface has multiple features to monitor/update segmentation and CT labeling, these steps were not considered and evaluated in this study. Thus, to overcome these limitations and improve ML model or image marker accuracy, in our future studies, we will ask experienced clinicians to validate/correct the possible errors in automated markings to improve region segmentation results, thereby, improving the performance of prediction models.

In conclusion, despite several limitations, this is a valid pilot study that has clearly demonstrated feasibility and potential advantages of developing new quantitative image markers to predict the prognosis of aSAH patients corresponding to various clinical measures used in current clinical practice. The initial results indicate the significance of both the admission (first) and day 10-14 (last) CT scans in their predictive capability of assessing both short-term and long-term clinical measures. Based on the foundation built in this study, more research efforts can follow to further optimize and validate the novel quantitative image markers by conducting prospective clinical studies in the future.

6 An Interactive Computer-Aided Detection Software Tool for Quantitative Estimation of Intracerebral Hemorrhage

6.1 Introduction

Intracerebral hemorrhage (ICH) is the most common type of intracranial bleeding that occurs with stroke. Among these ICH patients, more than one-third die within a month, and only about twenty percent will regain functional independence after treatment [123]. The volume of ICH is a vital variable used to calculate ICH-score, which is a well-known predictor of ICH prognosis [124]. Manual delineation of blood markings to compute the accurate volumes is tedious, time-consuming, and not practical in clinical settings. Currently, ABC/2 is the most used approximation method to calculate hemorrhagic volume [125].

During ICH, a cerebral bleed can occur within the brain tissue (intraparenchymal hemorrhage (IPH)) or ventricles (intraventricular hemorrhage (IVH)). Prior studies have investigated various qualitative methods to derive volumetric analysis of intraparenchymal hemorrhage (IPH), but IVH was not included. Based on our previous experiences in developing CAD schemes from neuroimaging [60], [59], [126], we believe a more accurate, fast, and reliable way for estimation of ICH volume will be very valuable for rapid prognosis and appropriate treatment, especially for those who are potentially fit for emergency surgical intervention. Thus, the purpose of this study is to develop a semi-automated interactive computer-aided detection (ICAD) segmentation tool for quantitative estimation of ICH volume.

6.2 Materials and methods

6.2.1 Image Dataset:

A non-contrast CT image dataset of 200 patients out of the 411 patients admitted with ICH diagnosis between the years 2012-2015 was retrospectively collected for this study. The imaging database was assembled from the existing clinical database from the Department of Neurology at the University of Oklahoma Health Sciences Center (OUHSC). During the period of hospitalization, multiple CT scans of the head were acquired from time to time for analyzing the patient's condition. For this study, we only used the initial admission scan before treatment for the quantification of ICH.

6.2.2 3D Brain Segmentation and CT Labeling:

We proposed a novel segmentation algorithm using a mapping technique in this CAD scheme. For each 3D Dicom CT sequence, the ICAD scheme performs an initial pre-processing step involving rough brain segmentation using thresholding is performed to identify the largest single-connected region in each of the 2D-image slices. Then based on segmented brain volumes, the slice with the largest volume is used as a reference marker to begin the proposed segmentation scheme. The fine-tuning begins at the reference marker slice as the intracranial brain volume in this slice is usually the most significant, single connected component and completely enclosed within the skull, facilitating accurate segmentation. Then, fine-tuning segmentation continues towards either left or right, applying a consecutive mapping technique. This mapping-based segmentation technique uses the prior slices segmentation result as guidance to identify multiple connected brain regions and act as limiting boundaries to avoid segmentation leakages. To identify multiple-connected regions, each region overlap is examined with prior segmentation results for its inclusion. Additionally, the limiting boundary criterion performs morphological dilation on

prior segmentation results to define the current slice boundary limit avoiding segmentation leakages. In summary, the above-mentioned process begins at identified reference marker slice and continues in both directions covering all the image slices.

After successful segmentation of the intracranial brain region, each CT voxel is labeled into three regions, namely, Cerebrospinal fluid (CSF), Tissue (White Matter, Grey Matter), and Blood using the Hounsfield Unit (HU) of the brain. The preset threshold (TH) values were TH1: 15, TH2: 45, and TH3: 80. The protocol set for this CT labeling is as follows: (1) CSF $\leq$ TH1, (2) TH1 $<$ Tissue $\leq$ TH2, (3) TH2 $<$ Blood $\leq$ TH3. After CT Labelling, each image slice is examined to identify the presence of IPH and IVH regions, followed by providing a contour enclosing the IPH and IVH. The ICAD scheme then labels blood within the boundary to IPH or IVH markings, respectively. Additionally, the combination of IPH and IVH labels is also categorized as the ICH.

Next, these markings were visually examined by a medical resident. In case of irregularities in these markings, the ICAD tool allows the residents to perform correction by drawing a new marking, thereby resulting in an accurate segmentation of IPH and IVH markings. Figure 6-1 illustrates the proposed ICAD segmentation tool for semi-automated ICH markings with multiple visual functionalities.

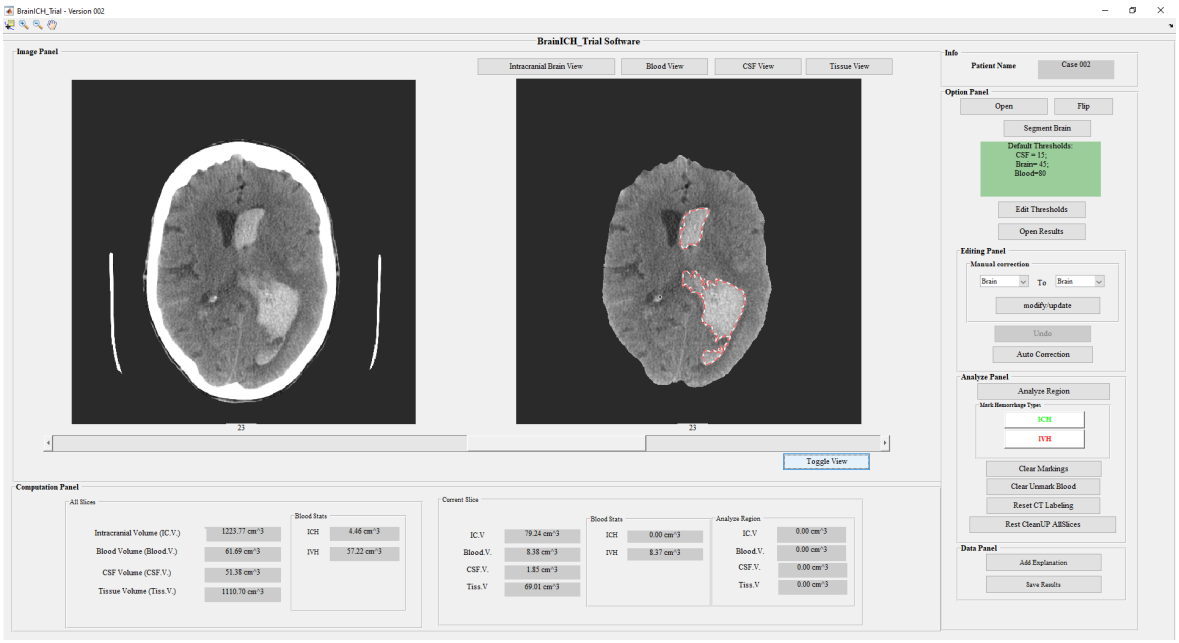


Figure 6-1: Proposed ICAD segmentation tool for semi-automated ICH markings.

6.2.3 Volumetric Analysis and DICE Similarity:

For each CT Label (i.e., IPH, IVH, and ICH), the CAD scheme computes image features. First, for each slice, the number of voxels belonging to the label are counted (N_{LABEL}), then voxel length V_L, voxel breadth V_B, and slice thickness S_T are used to calculate the respective label volume of each slice, as shown in equation (6.1). Next, a total case-based label volume is calculated as shown in equation (6.2). For example, to calculate the total amount of ICH, the number of ICH voxels in each slice to compute the total ICH volume as shown in equation (6.3). Similarly, the total volumes of V_{IPH}, *and* V_{IVH} were also computed. Additionally, the total intracranial brain volume enclosed within skull is also calculated as shown in equation (6.4). Additionally, for each case, the maximum 2D diameter (max_2D) of the largest CT label (i.e., IPH, IVH, and ICH) was also computed to capture information similar to the traditional ABC/2 method.

$$V_{LABEL}(slice) = N_{LABEL}(\text{slice}) \times V_L \times V_B \times S_T\,(\text{slice}) \tag{6.1}$$

$$\mathrm{V_{LABEL}} = \sum_{slice_k=m}^{n} \mathrm{V_{LABEL}}(slice_k) \tag{6.2}$$

$$\mathrm{V}_{ICH}(slice) = N_{ICH}(\mathrm{slice}) \times \mathrm{V_L} \times \mathrm{V_B} \times \mathrm{S_T}\ (\mathrm{slice}); \quad \mathrm{V}_{ICH} = \sum_{slice_k=m}^{n} \mathrm{V}_{ICH}(slice_k) \tag{6.3}$$

$$\mathrm{V_{Brain}}(slice) = N_{Brain}(\mathrm{slice}) \times \mathrm{V_L} \times \mathrm{V_B} \times \mathrm{S_T}\ (\mathrm{slice}); \quad \mathrm{V_{Brain}} = \sum_{slice_k=m}^{n} \mathrm{V_{Brian}}(slice_k) \tag{6.4}$$

Finally, the contribution of the semi-automated segmentation of the ICAD scheme towards accurate segmentation of ICH is made using the Dice similarity coefficient (DSC). To calculate DSC between the semi-automated markings (A) and the final accurate resident markings (B) an overlap between two markings is calculated as shown in equation (6.5).

$$dice(A,B) = 2 * |intersection(A,B)| / (|A| + |B|) \tag{6.5}$$

6.3 Results

For each case, the first segmentation of intracranial brain volume is performed before CT labeling and ICH markings. Figure 6- 2 illustrates the segmentation results of the ICAD scheme for ICH marking in a sample case. Figure 6- 2 only shows the slices consisting of the ICH markings here, where IPH and IVH are outlined using green and red colors, respectively.

Table 6-1 summarizes information of various volumetric parameters and maximum 2D diameter of CT labels (i.e., IPH, IVH, ICH, and intracranial brain). For each parameter, both median and interquartile range (IQR) are represented for all cases. Additionally, the volume-related parameters were represented in cubic.mm, and the diameter parameter is represented using mm.

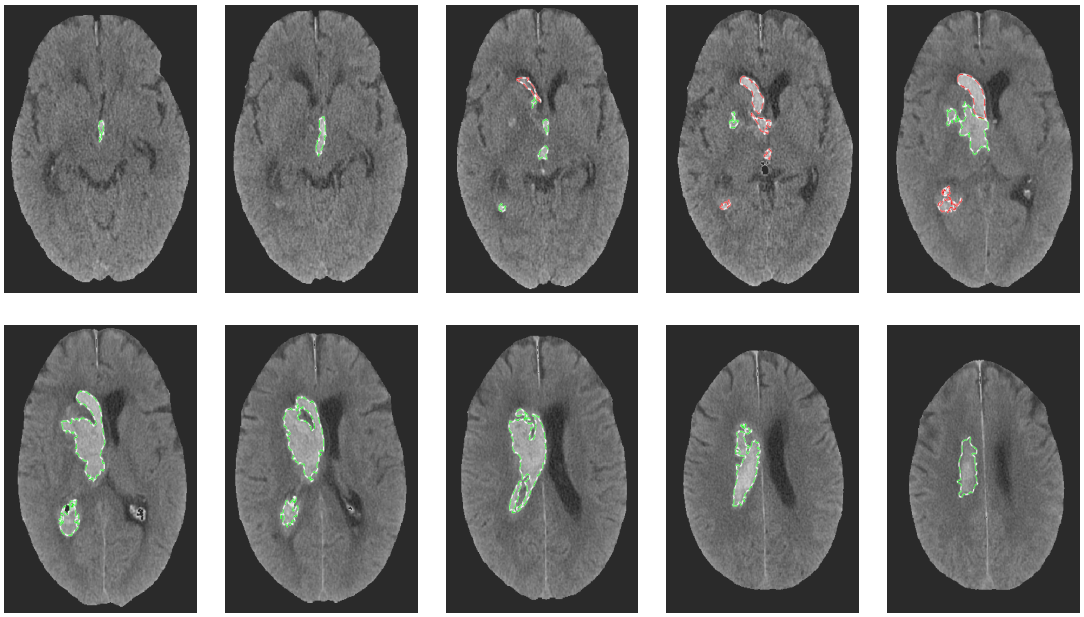


Figure 6- 2: A sample case illustration of ICAD scheme with ICH markings.

Table 6-1: Summary of various parameters of CT labels.

Label Type	Volumetric parameter (cubic.mm)		Max_2D parameter (mm)	
	Median	**IQR**	**Median**	**IQR**
Intracranial Brain	1279.7205	[1187.01, 1379.23]	-	-
ICH	37.8525	[13.6626, 70.7912]	54.5683	[38.0556, 71.7985]
IPH	24.5380	[7.3519, 54.2804]	49.7232	[31.9088, 63.8661]
IVH	10.8981	[2.3493, 36.0557]	39.1904	[23.3227, 64.2572]

Table 6-2 illustrates summary statistics of DSC between the semi-automated and resident markings for each of the CT labels (i.e., IPH, IVH, and ICH). Additionally, these measures were also categorized between the two observers who independently performed the markings for 100 cases.

Table 6-2: Summary statistics of DSC parameters both combined and separated between two observers.

Label Type	DICE: Combined		DICE: Observer-1		DICE: Observer-2	
	Median	**IQR**	**Median**	**IQR**	**Median**	**IQR**
ICH	0.9655	[0.9194, 0.9872]	0.9721	[0.9312, 0.9902]	0.9613	[0.9127, 0.9840]
IPH	0.9733	[0.9358, 0.9934]	0.9754	[0.9436, 0.9932]	0.9653	[0.9245, 0.9945]
IVH	0.9156	[0.8313, 0.9786]	0.8999	[0.8030, 0.9609]	0.9300	[0.8595, 0.9923]

6.4 Discussion

This study develops and evaluates a novel semi-automated and interactive computer-aided detection (ICAD) segmentation tool for quantitative estimation of ICH volume categorized by IPH and IVH markings. The study has several unique characteristics and contributions. First, the scheme has the automatic capability to perform various tasks, including (i) segmentation of intracranial brain volume, (ii) CT labeling if each voxel into three defined brain regions (i.e., CSF, Tissue, and Blood), (iii) markings of blood region into either IPH or IVH based on provided contours, (iv) computing various volumetric parameters. Additionally, the scheme allows for visual inspection and modifications of ICH markings followed by computing dice similarity between semi-automated and the final accurate segmentation of markings. Second, the semi-automated segmentation results were satisfactory, with the median DSC values ranging between 0.92 to 0.97 for ICH markings, as shown in Table 6-2. Thus, the time required to correct these markings to obtain accurate segmentation results using the ICAD tool is significantly less as compared to a fully manual segmentation scheme. Third, both the observers had comparable confidence in using the ICAD tool for segmentation with median DSC values of 0.97 and 0.96 (0.01 difference) for combined ICH markings. Thus, the initial ICH markings provided by the

ICAD act as a reference for marking and can help reduce inter-reader variability between the observers.

Despite the encouraging results, we recognize that this study has several limitations. First, the ICAD is a semi-automated segmentation scheme needing the user to draw contour enclosing ICH region to identify the ICH. In the future, we will use these segmentation markings as input for a deep-learning architecture to perform fully automated segmentation of both IPH and IVH [25]. Second, only one observer looked at each case independently; in the future, two observers will examine a subset of the same sample of cases to observe the inter-reader variability. Third, the current observers used the ICAD tool segmentation results as a reference, thereby being biased. In the future, we will ask a new observer to perform all the markings completely manually without the ICAD tool to further deduce the advantages of the ICAD scheme in terms of segmentation accuracy and time spent on each case. Furthermore, the application of advanced machine learning models with a multitude of imaging features used in medical analysis needs to be further investigated [101], [105], [127], [128]. In conclusion, this study has demonstrated the potential of our ICAD segmentation tool for marking and quantification of ICH using brain CT images of patients.

7 Comparison of Performance in Breast Lesions Classification Using Radiomics and Deep Transfer Learning: An Assessment Study

7.1 Introduction

Full-field digital mammography (FFDM) is the most common and widely accepted clinical imaging modality for breast cancer screening in the general population. However, FFDM has a relatively lower sensitivity and specificity due to two-dimensional projection imaging. Thus, it is challenging to develop computer-aided detection and diagnosis (CAD) to assist radiologists in detecting suspicious lesions and classifying between malignant and benign lesions. Currently, computer-aided detection (CADe) schemes have been routinely implemented in the clinical practice, while computer-aided diagnosis (CADx) schemes have not been accepted in clinical practice. In previous CAD studies, two technologies have been widely used to extract and compute image features for lesion classification.

First, traditional feature engineering to capture radiomic information is popular and well accepted in developing CAD schemes of medical images. Based on the radiomics concept, CAD schemes can extract a vast number of handcrafted features specific to understand the underlying phenomenon of suspicious breast lesions. These radiomic features can be obtained from a wide range of topics covering shape, density, texture patterns, frequency domain features etc. However, this higher number of initial feature dimensions comes with an inherent challenge of possible overfitting (curse of dimensionality), redundant information between features due to high correlation. Thus, it is important to take precautionary measures to reduce feature numbers or dimensionality. The optimal features can be obtained from either feature selection or reduction techniques. Additionally, the radiomics approach often faces the challenge of accurately segmenting subtle lesions if needed before feature engineering of local regions.

Second, in recent years, the interest in extracting automated features using the deep transfer learning method is emerging. Transfer learning exploits the phenomenon of learning global features independent of image types to initialize the network weights on larger and more commonly available images. Then, the pre-trained network can be finetuned using a small medical image dataset relative to the specific application task. However, in the medical image analysis, these 'black box' type, image-in, and prediction-out schemes are not easy to gain the trust of medical professionals.

Since in previous studies, CAD schemes are separately developed using either handcrafted radiomics features or deep transfer learning model generated automated features using different and relatively small image datasets. Thus, it is very difficult to compare the performance of these two types of image features to achieve better performance. As a result, the advantages and/or potential limitations of CAD schemes trained using radiomics and automated features have not been well investigated to date. In order to address this issue, we conduct a new study to explore the association/correlation between the traditional radiomics feature-based CADs and deep learning framework-based CAD scheme in classifying between malignant and benign breast lesions using a relatively large and diverse image dataset. Additionally, we also investigate whether integration of these two types of features further improves performance in lesion classification.

7.2 Materials and methods

7.2.1 Image Dataset

A fully anonymous and retrospective database consisting of full-field digital mammograms (FFDMs) was assembled for this study. The dataset is heterogeneous and consists of 2,778 FFDM images from craniocaudal (CC) and mediolateral oblique (MLO) views. The center location

belonging to suspicious lesion (soft tissue mass) in each image was marked by the radiologist. Based on the biopsy results, these images depict 1,452 malignant and 1,326 benign masses.

The study primarily consists of two main phases, (i) a traditional image analysis phase with details involving the handcrafted features (ii) a deep learning architecture adjusted and finetuned for generating probabilities to classify between benign and malignant classes. More details regarding these two phases, along with model evaluation settings, are explained in the following sections.

7.2.2 Image Processing and Traditional Feature Engineering

During the traditional image processing phase, we first examined all the cases to identify the ideal size of a rectangular window centered around the radiologists marking enclosing the lesion region. We observed the optimal window size to be 150×150, covering all types of lesions in the dataset. Then, we cropped the fixed-size image patches centered with the reference markings for each case. Necessary steps were taken to zero pad the edges or corners if the central region is along the boundary. Additionally, a relatively small subsample of cases consisting of chest wall regions within the patch was automatically segmented out. Then, we performed an adaptive thresholding-based segmentation with seed selected at the center. The segmentation results are satisfactory, and only a small subset (<5%) needed a manual adjustment of the segmentation boundary.

Next, a total of 235 traditional handcrafted image features covering a variety of radiomic information representing lesion characteristics such as shape, density, boundary contrast, texture patterns, and wavelets were computed. The lesion-specific features explain the local patterns like shape, density distribution within and around the boundary region. In contrast, global image features capture the total image patch's texture, density patterns, and frequency domain

information. More detailed information regarding these features can be found in our previous studies [1, 2].

7.2.3 Deep Learning Framework Settings

We used the popular image classification architecture of pre-trained residual net architecture (ResNet50) for the deep learning phase with weights tuned for the ImageNet dataset consisting of 1,000 classes. The final fully connected (*FC*) layer used for prediction was adjusted to categorize two classes (benign or malignant lesions). Then, we feed image patches of size 150×150 into the deep learning architecture and required transformations such as resizing (224×224×3: Height × Width × Depth), normalization of the mean and standard deviation of each channel were performed on the fly. We used the same grayscale FFDM image patch repeated for the three channels for the depth. Additionally, a minimal augmentation step (involving random centered crop, random horizontal, and random vertical flip with p=0.5) was added to introduce slight variation of a sample image for different epochs during the training phase. Due to the nature of medical images, a simple feature extractor type training involving freezing of all unchanged layers and updating only the weights and biases of the modified last FC layer did not yield good results. Thus, in this study, we optimized the weights of all layers during the training.

Given the limitation of our dataset size relative to the computer vision field, we maximize the training and consider the time required for this network-tuning; currently, we used 10-fold cross-validation (CV). During each fold, the data is split randomly into training (90%) and testing (10%) without repetition between them, and each sample case is only used once in the test phase. We investigated various batch sizes (i.e., 4,8,16, etc.) and found that a batch size of 4 works well for our analysis. Additionally, we selected Adam optimizer with an initial learning rate (*lr*) of 1e-4 at the beginning of each cross fold. We updated the learning rate scheduler with an exponential decay

function with a gamma value of 0.4 after each epoch. After each epoch, the network was evaluated to monitor training and validation loss during the training process, thereby deciding the stopping criterion. We noticed that by ten epochs, the network is saturated, and any further training resulted in overfitting. Thus, we only trained the network for ten epochs during each cross fold. During the validation phase, the network is loaded in evaluation mode, and a forward pass of data is done to collect both classification labels and probabilities. In summary, we used a 10-fold CV with ten epochs per each fold; at the end of each training fold, the test data was evaluated on the network to record both classification labels and the associated probabilities.

7.2.4 Model Building and Performance Evaluation:

We build and test several models to classify suspicious breast masses into two classes. Specifically, we investigated: (i) using only standard radiomic features, (ii) using probability score from ResNet50, (iii) integrated models with a combination of radiomic and ResNet50 models. More details regarding the models are provided in **Error! Reference source not found.**. In model I, the initial feature dimension of size 235 is reduced using PCA with a variance rate of 0.99, and then an SVM classifier was implemented. In model II, a simple classification based on a prediction probability of ResNet50 was performed. Whereas, during model III, multiple combinations of the above two were conducted using the output scores of models I and II, including III.1 using two scores from the model I and II considered as features to build an SVM classifier, III.2 using a simple weighted average of classification scores generated by models I and II, III.3 using a minimum score of models I and II, and III.4 using a maximum score of models I and II. Additionally, we also investigated the classification performance of three subgroups of traditional radiomic features (a. shape + density, b. wavelets, c. texture groups) with the integration of PCA into their respective SVM classification learner. The classification scores of each model were

named using '*S*' followed by the subscript of the model number. For instance, the model I output score is S1, and a weighted average model built using a combination of S_1 and S_2 is termed as $S_{3.2}$.

To evaluate the performance of each model, we used two steps. First, a receiver operating characteristic curve (ROC) is constructed from the classification scores. The area under the ROC curve (AUC) is computed and used as an index to evaluate and compare the performance of each model to classify between two classes. Second, we apply an operating threshold on the classification scores (T = 0.5) to divide all testing cases into two classes (score ≤ 0.5: 'Benign'; score > 0.5: 'Malignant'). Figure 7-1 shows a detailed flow chart explaining each step of the proposed CAD scheme.

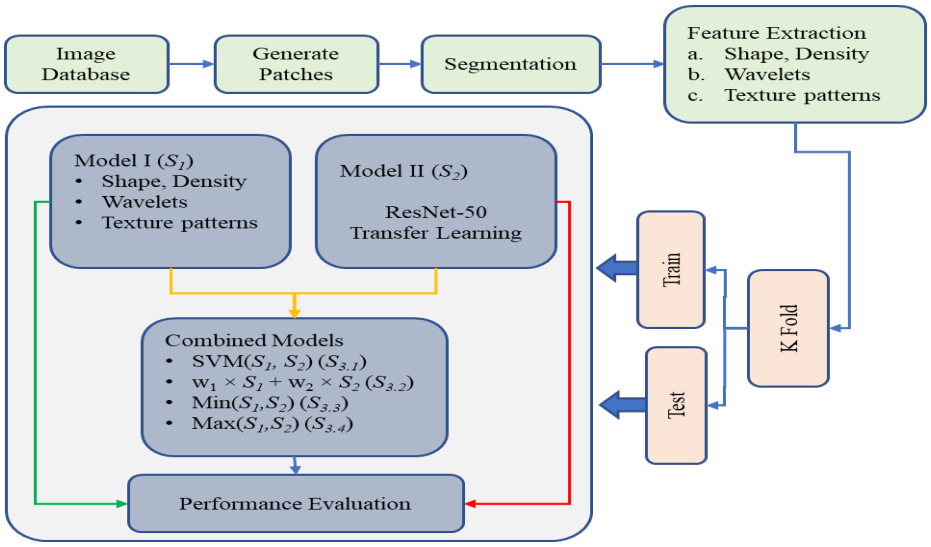


Figure 7-1: A detailed flow diagram of each step of the proposed CAD scheme.

7.3 Results

Figure 7-2 shows sample images in the database with an overlay of the segmentation results. The cases with segmentation overlay with red or green color are malignant or benign cases, respectively. Additionally, in some image patches, we can also notice that zero paddings are performed whenever needed (lesion at the edge or corner inside the original image). We can also

notice that the density distribution of lesions consists of both solid and diffused samples. The diffused or hidden lesions segmentation is challenging to segment and/or analyze and may not represent the underlying lesion image marker.

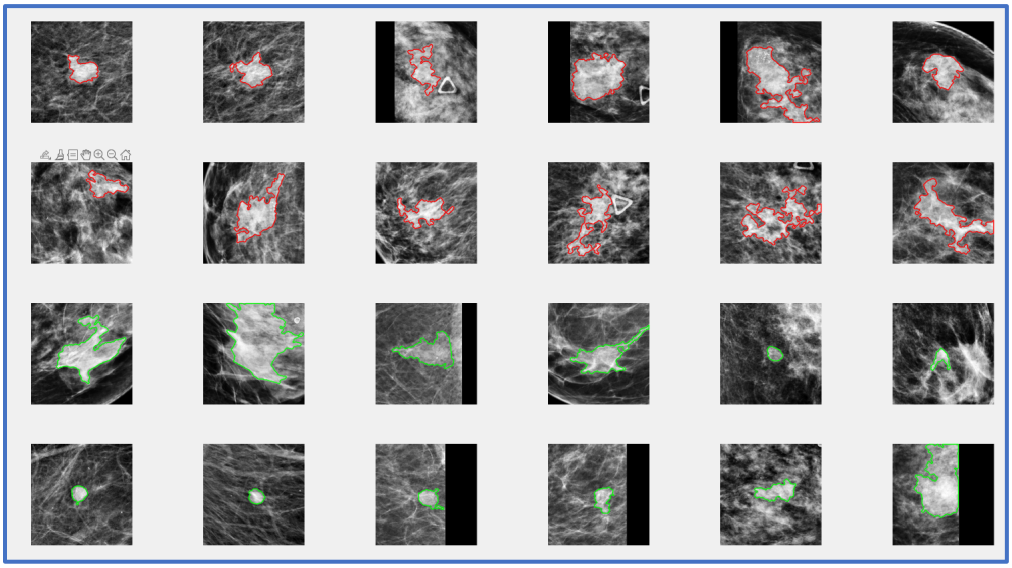


Figure 7-2: Sample case image patches with segmentation overlay (Red: Malignant; Green: Benign).

We first performed an independent analysis of each subgroup in model I during the performance evaluation stage before the models summarized in **Error! Reference source not found.**. These models were also evaluated using PCA integration with an SVM classifier during each cross fold. The first model was built using a subgroup of features, including shape and density-related features. It included a total of 50 features, and the distribution of ACCs was 65.68±0.02. Similarly, the models built using wavelet and texture pattern features independently resulted in an ACC distribution of 64.39±0.04 and 61.94±0.02, respectively. These traditional radiomic-based models' performance was the lowest as expected as we were only observing the classification capabilities separately. Whereas, when a combined model of these three subgroups was performed (model I), we observed an increase in the performance metrics (ACC, AUC) as shown in Table 7-1, indicating that the combination of these types adds new information for classification model to learn new information. Next, the ResNet50 (model II) performed

significantly better than the model I in terms of both ACC (77.31±2.65) and AUC (0.85±0.02). The trendline depicting the change in the improvement of performance distribution in terms of ACC for each fold per epoch using the ResNet50 is shown in Figure 7-3.

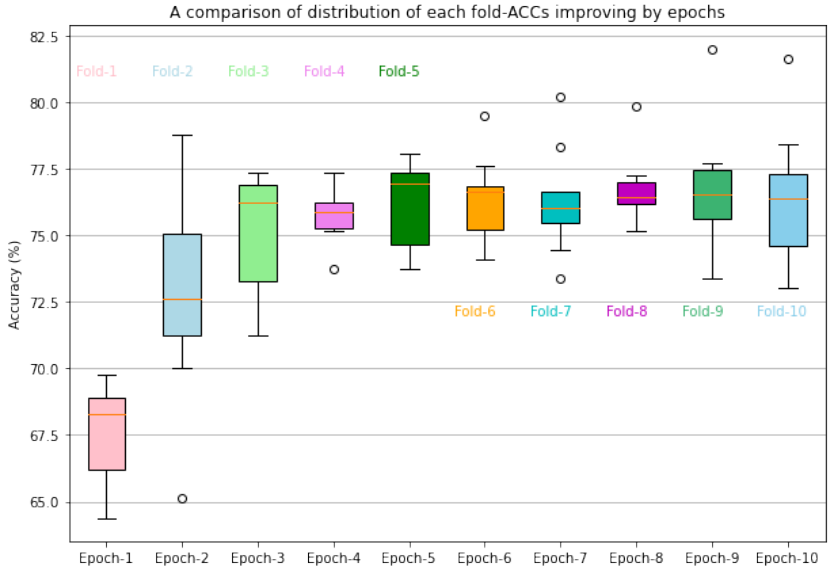


Figure 7-3: Trendline depicting the change in ACC distribution for each fold per epoch using ResNet50.

Next, we used four different combinations to observe the performance improvement combining models I and II scores. In models III.1 and III.2 built using SVM and weighted averages of S_1 and S_2, we noticed that the performance metrics are very similar to that of ResNet50. This indicates that both traditional and deep learning features converge at the end towards classification prediction and have a high correlation. Additionally, a negative effect on performance was observed when using either min- or max-based simple classification models. A more detailed comparison of the distribution of performance metrics for each model is shown in Figure 7-4. In terms of both ACC and AUC from these results, we clearly notice that ResNet50 or a combination of ResNet50 with traditional radiomic features yield a similar performance with a high association.

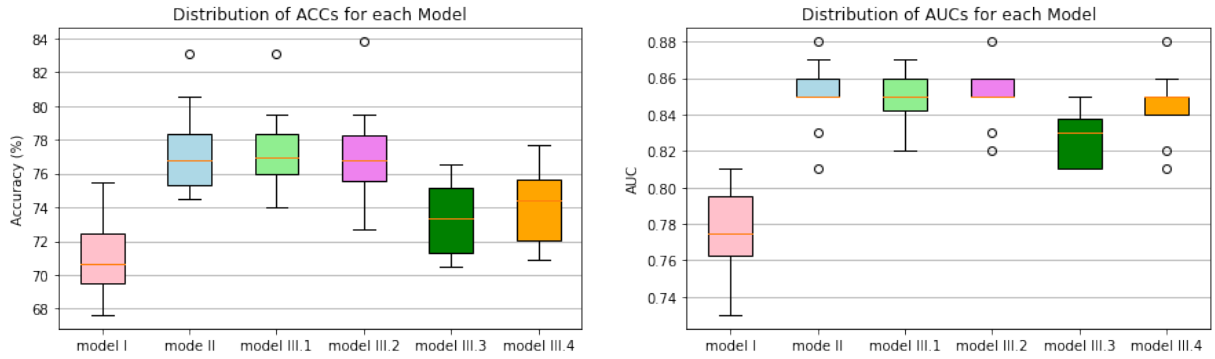


Figure 7-4: Comparison between distribution of (a) accuracies and (b) AUCs for each model.

Table 7-1: Summary of details for each model, including feature description and performance metrics.

Model (output score)	Feature description	AUC	ACC (%)
Model I (S_1)	a. shape + density, b. wavelets, c. texture groups	0.77 ± 0.02	71.23 ± 2.44
Model II (S_2)	classification probability of ResNet50	0.85 ± 0.02	77.31 ± 2.65
Model III.1 ($S_{3.1}$)	SVM (S_1, S_2)	0.85 ± 0.01	77.42 ± 2.47
Model III.2 ($S_{3.2}$)	$W_1 \times S_1 + W_2 \times S_2$	0.85 ± 0.01	77.31 ± 2.83
Model III.3 ($S_{3.3}$)	Min (S_1, S_2)	0.83 ± 0.02	73.35 ± 2.17
Model III.4 ($S_{3.4}$)	Max (S_1, S_2)	0.85 ± 0.02	74.07 ± 2.24

7.4 Discussion

In this study, we investigate the association between the traditional radiomics-based CAD features and the deep learning framework (ResNet50) in the classification of breast masses. This study generates several new and interesting observations, namely, first, the performance of subgroups of radiomic features is low when evaluated separately. Whereas, when combined, new information from each subgroup contribute additional information for the classification learner to improve the performance (ACC) significantly from a range for individual models (65.68±0.02, 64.39±0.04, 61.94±-0.02) to the combined model I (71.23±2.44).

Second, training a complex deep learning framework (ResNet) by freezing all input layers is not ideal given the contrast between ImageNet data and complex structures in the breast region.

However, a significant improvement in results compared to the model I is achieved by retraining a transfer learning model to update weights of all the layers in the network. Initializing the deep learning framework with weights from pre-trained ImageNet and customizing for a binary classification task (i.e., classifying between malignant and benign breast lesions in this study) works well. This step of careful customization and training all the layers for certain epochs is essential for optimally applying the deep transfer learning network to learn the parameters used in CAD of medical images.

Third, when we combine the scores from models I and II, we observe no significant change in the improvement of the performance as compared to model II. Nevertheless, this supports the theory that both optimized radiomic features and deep learning features have a high degree of correlation. This level of closeness/association between two classification scores represents that even though no specific attention is taken in the deep learning framework after careful selection of features, they still capture the underlying phenomenon of breast images similar to that of radiomic features. Thus, even a simple weighted average base classifier model built using a combination of scores S_1 and S_2 yielded a consistent performance similar to that of ResNet50.

Despite the encouraging observations in our results, we recognize some limitations in our study. First, even though we used a wide range of radiomic features (shape, density, texture, wavelets) for our model I, there are numerous more combinations to analyze the traditional feature-based model. Second, we only used the standard ResNet50 by modifying the last *FC* layer to examine and utilize the full potential of transfer learning. We need to validate this phenomenon on other state-of-the-art deep learning frameworks in the future. Third, we were limited to using 10-fold cross-validation given the constraints; in the future, we will try to use either more folds or

obtain a separate new test dataset to validate the classification performance of these transfer learning models.

In conclusion, this study observes and compares traditional radiomic feature-based CADs and deep learning framework-based CAD to classify breast masses. Additionally, we also observed a high degree of correlation between the classification scores of two types of CAD models, representing that both preserve/capture similar information irrespective of the discrepancies in both approaches. Thus, although deep transfer learning is widely considered a “black-box” type study with a high degree of difficulty for human users to understand its learning or decision-making logic or reasoning, the automated features provide high discriminatory information or power than traditional radiomics features. A furthermore comprehensive analysis covering both radiomic and deep learning architectures needs to be investigated to validate these observations.

8 Conclusion and Future Work

8.1 Summary

During the last several decades, research interests have grown in developing CAD schemes of medical images in commercial companies and research institutions. Many commercialized CADs are now available and utilized in clinical research and practice. Some commercialized CAD systems are clinically accepted as "second readers" to assist radiologists in interpreting medical images. However, the application of CAD in clinical practice is still limited, and more development effort and progress are needed. For this purpose, the proper supervision must be taken in multiple stages to develop a robust and reliable CAD scheme, which include (i) elimination of artifacts and improving the quality of images, (ii) Accurate detection and segmentation of the region of interest, (iii) radiomic feature computation and their optimization, and (iv) fine-tuning parameters of the machine learning model to address the underlying medical application objectives. Recently, this trend of CAD development and implementation has been accelerated due to many catalysts like (i) improved imaging technologies, (ii) growth in computational processing speeds, (iii) evolution of the concept of radiomics which depicts phenotype features that are highly associated with genomic and radiologic markers, (iv) advancements and application of deep learning architectures, (v) more research interest in the field of machine learning focusing on medical imaging informatics, etc. Despite encouraging results of developing various CAD systems in medicine, it is still an emerging field and requires more research penetration in many fields.

Many clinical studies involving the interpretation of medical images can be assisted and/or improved by identifying novel radiographic image markers that capture clinically observed patterns. Many research studies have shown high inter-and intra-reader variability in assessing medical images. Additionally, this manual interpretation is time-consuming, inconsistent, and

highly dependent on the reader's skill. Additionally, the role of radiomics has been proven in assessing radiographic images (i.e., CT and MRI) depicting useful image phenotype features that are highly associated with genomic biomarkers [95] and have the potential to predict disease prognosis [96]. Thus, developing and integrating such study-specific radiomic engineered CADs with current evaluation tools used by radiologists/ experts will be useful to serve as a secondary reader and address some of the existing limitations.

In the previous chapters of my dissertation, I presented and discussed several new approaches or solutions to address several key aspects of CAD development. The contributions in these studies ranged from developing novel image markers, segmentation architectures, and building interactive visually aided image analysis software tools to build problem-specific applications that are robust, reliable, and easily interpretable in assessing various medical imaging problems.

In chapter 3, I successfully validate my hypothesis of using the two CEDM images offering complementary information to improve CAD performance. For this purpose, a new and optimal approach is applied to develop a fully automated CAD scheme of CEDM images to classify suspicious breast lesions. One crucial observation in this study is that DES images constructed using two images (LE and HE) obtained at two x-ray energy levels can eliminate the overlapping effect of FGT and better highlight the tumors. The study demonstrates that both DES and LE images contribute complementary information that can be used to improve segmentation (DES) and extract more tumor-related heterogenous (LE) information to improve the overall lesion classification accuracy of the CAD scheme. The study is the first of its kind to develop an automated CAD scheme using CEDM images. It has established a solid foundation to continue developing and optimizing CAD schemes using CEDM images to classify breast lesions in the future.

In Chapters 4 and 5, I present several unique quantitative image markers extracted from brain CT images enabling to predict the prognosis of patients suffering from different types of stroke diseases. Specifically, in Chapter 4, I demonstrate the feasibility of extracting image markers from the CTP images at an early stage to predict the prognosis of AIS patients. Several unique image processing algorithms are proposed and embedded into building this CAD scheme. For instance, although parameters like (i) the number of image sequences, (ii) the number of image slices per sequence, etc., are inconsistent between different brain CT scanning cases. The new CAD scheme automatically identifies and normalizes these parameters before performing the rest of the image processing and feature analysis. Next, segmentation of brain regions and identifying the blood flow patterns over time are also carefully designed to develop novel image markers representing the asymmetrical blood volumes. Additionally, I design and implement a unique interactive CAD that supports full customization and the ability to monitor all aspects of the image analysis pipeline to encourage experienced readers to observe and optimize the automated results if needed in their research or future clinical applications. The study demonstrates promising results when applying the proposed CAD scheme and ML model to predict AIS patients' prognosis with the diverse class distribution.

In Chapter 5, image markers representing clinically relevant information are also automatically developed to investigate its potential in predicting a wide range of clinical measures amongst aSAH patients. Like the CAD scheme developed in Chapter 4, very careful customization of an interactive CAD scheme is also successfully developed and tested to automatically segment and generate radiomics image markers representing the proportions of various subregions. The study results are very encouraging and support the clinical expectations. Namely, (i) the admission CT scan illustrates its potential in predicting the short-term clinical complications, and (ii) the

discharge CT scan shows significance in predicting the prognosis of long-term measures. The study provides scientific rationale or evidence to support conducting future prospective studies to use the admission results generated by the CAD scheme to take more patient-specific care and treatment, thereby impacting the overall long-term recovery in the aSAH patients.

In Chapter 6, I demonstrate another example to support translational clinical research activities by developing and implementing an ICAD scheme or tool of image segmentation and image marker quantification, which allows the clinical researchers to semi-automatically segment and quantify the ICH markings. To improve the efficacy of using this ICAD tool by the clinical researchers, the ICAD tool only requires minimum user input, such as providing a rough boundary enclosing the ICH region. CAD will perform identification and automatic segmentation of 3D ICH volumes. The utility of this new ICAD tool has been successfully validated by the experienced clinical researchers in their translational studies that have reported quite encouraging results to identify and extract new quantitative image markers, which highly associate with the patients' prognosis or clinical outcomes.

In chapter 7, I demonstrate a comparative study that successfully assesses and compares the two most popular types of CAD approaches using either traditional radiomics-based or deep learning such as CNN-based CAD models. We assembled a relatively large image dataset of nearly 2,800 mammograms to investigate the association/correlation between these two types of CAD schemes in evaluating their performance in classifying suspicious breast lesions. The study results show that both types of CAD schemes contain similar information in their ability to classify breast lesions. Additionally, the CNN-based CAD model performs significantly better than the traditional radiomics feature-based CAD model. No significant performance improvement achieved using various fusion techniques further supports my research hypothesis as presented in Chapter 2.

In summary, the research efforts made during my Ph.D. studies provided me with a great learning opportunity to explore, investigate, and contribute to developing and testing several new novel CAD schemes or ICAD tools for medical images. Additionally, I am also part of many collaborative CAD studies along with researchers in both medical imaging engineering and clinical research or applications fields. These research efforts have resulted in promising results, as reported in many research papers published in engineering and clinical journals and medical imaging conference proceedings, demonstrating the scientific significance and potential translational clinical applications.

8.1.1 Journal Articles

1. **Danala G**, Desai M, Ray B, et al. Applying Quantitative Radiographic Image Markers to Predict Clinical Complications After Aneurysmal Subarachnoid Hemorrhage: A Pilot Study. Annals of Biomedical Engineering. doi: 10.1007/s10439-022-02926-z.
2. **Danala G**, Ray B, Desai M, et al. Developing New Quantitative CT Image Markers to Predict Prognosis of Acute Ischemic Stroke Patients. Journal of X-Ray Science and Technology. 2022.doi: 10.3233/XST-221138
3. Nouh CD, Ray B, Xu C, Zheng B, **Danala G**, et al. Quantitative Analysis of Stress-Induced Hyperglycemia and Intracranial Blood Volumes for Predicting Mortality After Intracerebral Hemorrhage. Transl Stroke Res. 2022. doi:10.1007/s12975-022-00985-x.
4. Heidari M, Lakshmivarahan S, Mirniaharikandehei S, **Danala G**, et al. Applying a random projection algorithm to optimize machine learning model for breast lesion classification. IEEE Trans Biomed Eng. 2021; 68(9):2764-2775.
5. Mirniaharikandehei S, Heidari M, **Danala G**, Lakshmivarahan S, Zheng B. Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using CT images. Comput Methods Programs Biomed. 2021; 200:105937.
6. Porwal P, Pachade S, Kokare M, et al. Idrid: Diabetic retinopathy--segmentation and grading challenge. Med Image Anal. 2020; 59:101561.

7. Heidari M, Mirniaharikandehei S, Khuzani AZ, **Danala G**, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. Int J Med Inform. 2020; 144:104284.
8. O'Connor KP, Hathidara MY, **Danala G**, et al. Predicting Clinical Outcome After Mechanical Thrombectomy: The GADIS (Gender, Age, Diabetes Mellitus History, Infarct Volume, and Current Smoker) Score. World Neurosurg. 2020; 134:e1130-e1142.
9. Shujaat SD, Xu C, Wu D, Ray B, **Danala G**, et al. Subarachnoid Hemorrhage Early Brain Edema Score is Associated with Shunt Dependence and Worse Clinical Outcomes (5112). Neurology. 2020; 94(15 Supplement).
10. Mirniaharikandehei S, VanOsdol J, Heidari M, **Danala G**, et al. Developing a Quantitative Ultrasound Image Feature Analysis Scheme to Assess Tumor Treatment Efficacy Using a Mouse Model. Sci Rep. 2019; 9(1):1-10.
11. Zheng B, Qiu Y, Aghaei F, Mirniaharikandehei S, Heidari M, **Danala G**. Developing global image feature analysis models to predict cancer risk and prognosis. Vis Comput Ind Biomed art. 2019; 2(1):58-71.
12. Ray B, Ross SR, **Danala G**, et al. Systemic response of coated-platelet and peripheral blood inflammatory cell indices after aneurysmal subarachnoid hemorrhage and long-term clinical outcome. J Crit Care. 2019; 52:1-9.
13. **Danala G**, Patel B, Aghaei F, et al. Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms. Ann Biomed Eng. 2018; 46(9):1419-1431.
14. Heidari M, Khuzani AZ, Hollingsworth AB, **Danala G**, et al. Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. Phys Med Biol. 2018;63(3).
15. **Danala G,** Thai T, Gunderson CC, et al. Applying Quantitative CT Image Feature Analysis to Predict Response of Ovarian Cancer Patients to Chemotherapy. Acad Radiol. 2017; 24:1233-1239.

8.1.2 Articles on Conference Proceedings

1. **Danala G**, Maryada SK, Huong P, et al. Comparison of performance in breast lesions classification using radiomics and deep transfer learning: an assessment study, In: SPIE Medical Imaging. 2022.
2. **Danala G**, Mirniaharikandehei S, Maryada SK, et al. Developing interactive computer-aided detection tools to support translational clinical research, In: SPIE Medical Imaging. 2022.

3. Islam W, **Danala G**, Pham H, et al. Improving the performance of computer-aided classification of breast lesions using a new feature fusion method, In: SPIE Medical Imaging. 2022.
4. Maryada SK, Booker WL, **Danala G**, et al. Applying a novel two-stage deep-learning model to improve accuracy in detecting retinal fundus images, In: SPIE Medical Imaging. 2022.
5. Pham H, **Danala G**, Zheng B, et al. Identifying an optimal machine learning generated image marker to predict survival of gastric cancer patients, In: SPIE Medical Imaging. 2022.
6. **Danala G**, Desai M, Shoukat M, Asif A, Heidari M, Zheng B. Applying Quantitative Image Markers to Predict Clinical Measures after Aneurysmal Subarachnoid Hemorrhage. In: SPIE Medical Imaging. 2021; 115972Z.
7. **Danala G**, Desai M, Astani S, Rabbani B, Garcia JGO, Zheng B. An Interactive Computer-Aided Detection Software Tool for Quantitative Estimation of Intracerebral Hemorrhage. In: SPIE Medical Imaging. 2021; 1160005V.
8. Heidari M, Mirniaharikandehei S, Khuzani AZ, **Danala G**, Qiu Y, Zheng B. Detecting COVID-19 infected pneumonia from x-ray images using a deep learning model with image preprocessing algorithm. In: SPIE Medical Imaging. 2021; 115970V.
9. Mirniaharikandehei S, Heidari M, **Danala G**, Lakshmivarahan S, Zheng B. A novel feature reduction method to improve performance of machine learning model. In: SPIE Medical Imaging. 2021:1159726.
10. Heidari M, Mirniaharikandehei S, Khuzani AZ, **Danala G**, et al. An optimal machine learning model for breast lesion classification based on random projection algorithm for feature optimization. In: SPIE Medical Imaging. 2021; 116010L.
11. **Danala G**, Maryada SKR, Heidari M, Ray B, Desai M, Zheng B. A new interactive visual-aided decision-making supporting tool to predict severity of acute ischemic stroke. In: SPIE Medical Imaging. 2020; 113171V.
12. Heidari M, Mirniaharikandehei S, **Danala G**, Qiu Y, Zheng B. A new case-based CAD scheme using a hierarchical SSIM feature extraction method to classify between malignant and benign cases. In: SPIE Medical Imaging. 2020; 1131816.

13. **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. In: SPIE Medical Imaging. 2019; 109530M.
14. Heidari M, Hollingsworth A, Mirniaharikandehei S, **Danala G**, et al. Assessment of short-term breast cancer risk using a frequency domain correlation based imaging marker. In: SPIE Medical Imaging. 2019; 109541F.
15. Mirniaharikandehei S, Heidari M, **Danala G**, Qian W, Qiu Y, Zheng B. Association of computer-aided detection results and breast cancer risk. In: SPIE Medical Imaging. 2019; 109500I.
16. **Danala G**, Aghaei F, Heidari M, Wu T, Patel B. Computer-aided Classification of Breast Masses Using Contrast-Enhanced Digital Mammograms. In: SPIE Medical Imaging. 2018; 105750P.
17. Heidari M, Khuzani AZ, **Danala G**, Mirniaharikandehei S, Qian W, Zheng B. Applying a machine learning model using a locally preserving projection based feature regeneration algorithm to predict breast cancer risk. In: SPIE Medical Imaging. 2018; 105790T.
18. 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. In: SPIE Medical Imaging. 2018; 105753N.
19. Khuzani AZ, **Danala G**, Heidari M, et al. Applying a new unequally weighted feature fusion method to improve CAD performance of classifying breast lesions. In: SPIE Medical Imaging. 2018; 105752L.
20. 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. In: SPIE Medical Imaging. 2018; 105752O.
21. Heidari M, Khuzani AZ, **Danala G**, Qiu Y, Zheng B. Improving performance of breast cancer risk prediction using a new CAD-based region segmentation scheme. In: SPIE Medical Imaging. 2018; 105750P.
22. **Danala G**, Wang Y, Thai T, Gunderson C, Moxley K, Moore K, Mannel R, Liu H, Zheng B, Qiu Y. Apply radiomics approach for early stage prognostic evaluation of ovarian cancer patients: a preliminary study. In: SPIE Medical Imaging. 2017; 1013449.
23. **Danala G**, Wang Y, Thai T, Gunderson C, Moxley K, Moore K, Mannel R, Liu H, Cheng S, Zheng B, Qiu Y. Improving efficacy of metastatic tumor segmentation to facilitate early prediction of ovarian cancer patients' response to chemotherapy. In: SPIE Biophotonics and Immune Responses XII. 2017; 100650J.

8.2 Future Works

Despite extensive research efforts and progress made in the CAD field, researchers still face many challenges in developing robust CAD schemes for clinical applications. The role of radiomics to represent the tumor heterogeneity information needs to be further investigated on a larger and more diverse dataset of many diseases. Additionally, even though the state-of-art deep learning systems have demonstrated better performance in classification and segmentation tasks. These systems have a few limitations, namely, (i) evaluated on smaller datasets using augmentation technique, (ii) Black-box approach of image-in, and prediction-out schemes are not easy to gain the trust of medical professionals. Thus, there is a need to build CAD systems that can be visually interactive, allowing the clinical researchers to observe various stages of deep learning

Based on my experience during my Ph.D. study, I will continue my research to explore further techniques to improve the acceptance of CAD applications in clinical practice. For instance, I will further validate my CAD schemes' performance on larger datasets collected in our future prospective studies. In a few of my studies (i.e., chapters 4,5,6), I am currently working with clinical researchers to use these CAD systems and optimize them based on their recommendations. Additionally, deep learning systems can be used to automate various aspects of CAD systems. Thus, I will explore the role of such systems to act as feature extractors or segmentation tools in my current studies. Then, I will integrate them with existing ML systems embedded along with ICAD tools. Finally, any level of automation/ generalization of CAD schemes may not be enough to gain the trust of the radiologist in adapting them as assisting tools. Thus, I will continue to develop/improve study-specific customizable visual tools based on radiologists' recommendations in the CAD schemes.

By combining my previous works and additional new work in my Ph.D. research and studies, I can comprehensively investigate different machine learning algorithms' challenges as a frame of CAD systems and develop more robust algorithms. Since I am exploring the integration of CAD systems and deep learning, I believe the success of my research will provide a significant contribution to the development of CAD systems in the future in the medical imaging informatics area. Furthermore, my research achievements during my Ph.D. will significantly benefit the research work in my academic, professional development, and future career.

References

[1] A. Jemal, E. Ward, Y. Hao, and M. Thun, "Trends in the leading causes of death in the United States, 1970-2002," *J. Am. Med. Assoc.*, vol. 294, no. 10, pp. 1255–1259, 2005.

[2] R. T. Greenlee *et al.*, "Cancer Statistics , 2001," *CA Cancer J Clin*, vol. 51, no. 1, pp. 15–36, 2001.

[3] A. C. Society, "American Cancer Society. Cancer Facts & Figures 2020," *American Cancer Society*. pp. 1–52, 2020.

[4] A. S. Panayides *et al.*, "AI in Medical Imaging Informatics: Current Challenges and Future Directions," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 7, pp. 1837–1857, 2020.

[5] T. Kobayashi, X. W. Xu, H. MacMahon, C. E. Metz, and K. Doi, "Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs.," *Radiology*, vol. 199, no. 3, pp. 843–848, Jun. 1996.

[6] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *Br. J. Radiol.*, vol. 78, no. SPEC. ISS., 2005.

[7] B. Sahiner *et al.*, "Effect of CAD on Radiologists' Detection of Lung Nodules on Thoracic CT Scans: Analysis of an Observer Performance Study by Nodule Size," *Acad. Radiol.*, vol. 16, no. 12, pp. 1518–1530, 2009.

[8] M. C. B. Godoy *et al.*, "Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin- and thick-section CT," *Am. J. Roentgenol.*, vol. 200, no. 1, pp. 74–83, 2013.

[9] J. J. Fenton *et al.*, "Effectiveness of computer-aided detection in community mammography practice," *J. Natl. Cancer Inst.*, vol. 103, no. 15, pp. 1152–1161, 2011.

[10] L. Hadjiiski, B. Sahiner, and H. P. Chan, "Advances in computer-aided diagnosis for breast cancer," *Curr. Opin. Obstet. Gynecol.*, vol. 18, no. 1, pp. 64–70, 2006.

[11] M. Hecht, "Picture archiving and communication systems.," *Health Devices*, vol. 33, no. 8, pp. 265–290, 2004.

[12] R. W. Regenhardt *et al.*, "Blood pressure and penumbral sustenance in stroke from large vessel occlusion," *Front. Neurol.*, vol. 8, no. JUL, 2017.

[13] F. Sardanelli and F. Podo, "Breast MR imaging in women at high-risk of breast cancer. Is something changing in early breast cancer detection?," *Eur. Radiol.*, vol. 17, no. 4, pp. 873–887, Apr. 2007.

[14] D. Saslow *et al.*, "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography.," *CA. Cancer J. Clin.*, vol. 57, no. 2, pp. 75–89, 2007.

[15] F. Aghaei, M. Tan, A. B. Hollingsworth, and B. Zheng, "Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy," *J. Magn. Reson. Imaging*, vol. 44, no. 5, pp. 1099–1106, 2016.

[16] N. H. Eltonsy, G. D. Tourassi, A. S. Elmaghraby, and S. Member, "A Concentric Morphology Model for the Detection of Masses in Mammography," *IEEE Trans. Med. Imaging*, vol. 26, no. 6, pp. 880–889, 2007.

[17] P. Porwal *et al.*, "IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge," *Med. Image Anal.*, vol. 59, 2020.

[18] "GLOBAL REPORT ON DIABETES, World Heath Organization," *WHO*, 2016.

[19] L. Guariguata, D. R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J. E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035.," *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 137–149, Feb. 2014.

[20] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, "Global estimates of the prevalence of diabetes for 2010 and 2030.," *Diabetes Res. Clin. Pract.*, vol. 87, no. 1, pp. 4–14, Jan. 2010.

[21] Y. Zheng, M. He, and N. Congdon, "The worldwide epidemic of diabetic retinopathy.," *Indian J. Ophthalmol.*, vol. 60, no. 5, pp. 428–431, 2012.

[22] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs.," *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016.

[23] M. Haloi, "Improved Microaneurysm Detection using Deep Neural Networks," *CoRR*, vol. abs/1505.0, 2015.

[24] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, "Lesion detection and Grading of Diabetic Retinopathy via Two-stages Deep Convolutional Neural Networks," *CoRR*, vol. abs/1705.0, 2017.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.0, 2015.

[26] M. Abadi *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," in *12th {USENIX} Symposium*

on Operating Systems Design and Implementation ({OSDI} 16), 2016, pp. 265–283.
[27] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Histogram-Based Features Selection and Volume of Interest Ranking for Brain PET Image Classification.," *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, p. 2100212, 2018.
[28] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
[29] P. Morgado, M. Silveira, and J. S. Marques, "Diagnosis of Alzheimer's disease using 3D local binary patterns," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 1, no. 1, pp. 2–12, 2013.
[30] M. Heidari *et al.*, "Applying a random projection algorithm to optimize machine learning model for breast lesion classification," *IEEE Trans. Biomed. Eng.*, 2020.
[31] H. Müller, A. Rosset, A. Garcia, J.-P. Vallée, and A. Geissbuhler, "Informatics in radiology (infoRAD): benefits of content-based visual data access in radiology.," *Radiogr. a Rev. Publ. Radiol. Soc. North Am. Inc*, vol. 25, no. 3, pp. 849–858, 2005.
[32] M. O. Lam, T. Disney, D. S. Raicu, J. Furst, and D. S. Channin, "BRISC-an open source pulmonary nodule image retrieval framework.," *J. Digit. Imaging*, vol. 20 Suppl 1, no. Suppl 1, pp. 63–71, Nov. 2007.
[33] T. M. Lehmann *et al.*, "Automatic categorization of medical images for content-based retrieval and data mining.," *Comput. Med. imaging Graph. Off. J. Comput. Med. Imaging Soc.*, vol. 29, no. 2–3, pp. 143–155, 2005.
[34] Y. Liu, D. Zhang, G. Lu, and W. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognit.*, vol. 40, pp. 262–282, 2007.
[35] J. Vogel and B. SCHIELE, "Semantic Modeling of Natural Scenes for Content-Based Image Retrieval," *Int. J. Comput. Vis.*, vol. 72, no. 2, pp. 133–157, 2007.
[36] L. R. Long, S. Antani, T. M. Deserno, and G. R. Thoma, "Content-Based Image Retrieval in Medicine: Retrospective Assessment, State of the Art, and Future Directions," *Int J Heal. Inf Syst Inf.*, vol. 4, no. 1, pp. 1–16, 2010.
[37] and M. J. B. L. Zheng, A. W. Wetzel, J. Gilbertson, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, pp. 249–255, 2003.
[38] H. C. A. and M. N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Trans. Inf. Technol. Biomed*, vol. 16, pp. 758–769, 2012.
[39] X. Zhang, S. Member, W. Liu, and M. Dundar, "Towards Large-Scale Histopathological Image Analysis : Hashing-Based Image Retrieval," vol. 34, no. 2, pp. 496–506, 2015.
[40] J. Liu, S. Zhang, W. Liu, C. Deng, Y. Zheng, and D. N. Metaxas, "Scalable Mammogram Retrieval Using Composite Anchor Graph Hashing With Iterative Quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2450–2460, Nov. 2017.
[41] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: hashing-based image retrieval.," *IEEE Trans. Med. Imaging*, vol. 34, no. 2, pp. 496–506, Feb. 2015.
[42] Y. Xu, F. Shen, X. Xu, L. Gao, Y. Wang, and X. Tan, "Large-Scale Image Retrieval with Supervised Sparse Hashing," *Neurocomput.*, vol. 229, no. C, pp. 45–53, Mar. 2017.
[43] J. R. Geis, "Medical imaging informatics: How it improves radiology practice today," *J. Digit. Imaging*, vol. 20, no. 2, pp. 99–104, 2007.
[44] E. Bercovich and M. C. Javitt, "Medical Imaging: From Roentgen to the Digital Revolution, and Beyond," *Rambam Maimonides Med. J.*, vol. 9, no. 4, p. e0034, 2018.
[45] G. W. Albers *et al.*, "Thrombectomy for Stroke at 6 to 16 Hours with Selection by Perfusion Imaging.," *N. Engl. J. Med.*, vol. 378, no. 8, pp. 708–718, Feb. 2018.
[46] E. G. Hoeffner *et al.*, "Cerebral perfusion CT: technique and clinical applications.," *Radiology*, vol. 231, no. 3, pp. 632–644, Jun. 2004.
[47] W. A. Copen *et al.*, "In Acute Stroke, Can CT Perfusion-Derived Cerebral Blood Volume Maps Substitute for Diffusion-Weighted Imaging in Identifying the Ischemic Core?," *PLoS One*, vol. 10, no. 7, p. e0133566, 2015.
[48] B. Zhao *et al.*, "Exploring intra- and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals.," *Eur. J. Radiol.*, vol. 82, no. 6, pp. 959–968, Jun. 2013.
[49] S. C. Johnston, S. Selvin, and D. R. Gress, "The burden, trends, and demographics of mortality from subarachnoid hemorrhage.," *Neurology*, vol. 50, no. 5, pp. 1413–1418, May 1998.
[50] C. Melinosky, H. Kincaid, J. Claassen, G. Parikh, N. Badjatia, and N. A. Morris, "The Modified Fisher

Scale Lacks Interrater Reliability.," *Neurocrit. Care*, Nov. 2020.
[51] J. A. Frontera *et al.*, "Prediction of symptomatic vasospasm after subarachnoid hemorrhage: the modified fisher scale.," *Neurosurgery*, vol. 59, no. 1, pp. 21–27, Jul. 2006.
[52] B. Ray *et al.*, "Systemic response of coated-platelet and peripheral blood inflammatory cell indices after aneurysmal subarachnoid hemorrhage and long-term clinical outcome.," *J. Crit. Care*, vol. 52, pp. 1–9, Aug. 2019.
[53] H. A. Choi *et al.*, "Quantification of Cerebral Edema After Subarachnoid Hemorrhage," *Neurocrit. Care*, pp. 64–70, 2016.
[54] J. Y. Yuan *et al.*, "Automated Quantification of Reduced Sulcal Volume Identifies Early Brain Injury After Aneurysmal Subarachnoid Hemorrhage.," *Stroke*, vol. 52, no. 4, pp. 1380–1389, Apr. 2021.
[55] S.-H. Ahn *et al.*, "The Subarachnoid Hemorrhage Early Brain Edema Score Predicts Delayed Cerebral Ischemia and Clinical Outcomes," *Neurosurgery*, vol. 83, no. 1, p. 137—145, Jul. 2018.
[56] G. Danala, B. Ray, M. Desai, M. Heidari, S. K. R. Maryada, and B. Zheng, "Developing New Quantitative CT Image Markers to Predict Prognosis of Acute Ischemic Stroke Patients."
[57] G. Danala *et al.*, "Applying Quantitative Radiographic Image Markers to Predict Clinical Complications After Aneurysmal Subarachnoid Hemorrhage: A Pilot Study," *Ann. Biomed. Eng.*, vol. 50, no. 4, pp. 413–425, 2022.
[58] C. D. Nouh *et al.*, "Quantitative Analysis of Stress-Induced Hyperglycemia and Intracranial Blood Volumes for Predicting Mortality After Intracerebral Hemorrhage," *Transl. Stroke Res.*, 2022.
[59] S. D. Shujaat *et al.*, "Subarachnoid Hemorrhage Early Brain Edema Score is Associated with Shunt Dependence and Worse Clinical Outcomes (5112)," *Neurology*, vol. 94, no. 15 Supplement, 2020.
[60] K. P. O'Connor *et al.*, "Predicting Clinical Outcome After Mechanical Thrombectomy: The GADIS (Gender, Age, Diabetes Mellitus History, Infarct Volume, and Current Smoker) Score.," *World Neurosurg.*, vol. 134, pp. e1130–e1142, Feb. 2020.
[61] P. G. M. Peer, A. L. M. Verbeek, H. Straatman, J. H. C. L. Hendriks, and R. Holland, "Age-specific sensitivities of mammographic screening for breast cancer," *Breast Cancer Res.*, pp. 153–160, 1996.
[62] T. M. Mandelson, "Breast density as a predictor of breast cancer risk," *J. Natl. Cancer Inst.*, vol. 12, no. Suppl 3, pp. 1081–1087, 2000.
[63] M. B. Laya, E. B. Larson, S. H. Taplin, and E. White, "Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography," *J. Natl. Cancer Inst.*, vol. 88, no. 10, pp. 643–649, 1996.
[64] M. Kriege, C. T. M. Brekelmans, C. Boetes, and P. E. Besnard, "Efficacy of MRI and Mammography for Breast-Cancer Screening in Women with a Familial or Genetic Predisposition," *N. Engl. J. Med.*, vol. 351, pp. 427–437, 2004.
[65] P. A. Carney *et al.*, "Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography.," *Ann. Intern. Med.*, vol. 138, pp. 168–175, 2003.
[66] H. Rebecca A, K. Kerlikowske, C. I. Flowers, B. C. Yankaskas., Z. Weiwei, and D. L. Miglioretti., "Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography," *Ann. Intern. Med.*, vol. 155, no. 3, pp. 481–492, 2011.
[67] J. Brodersen and V. Siersma, "Long-Term Psychosocial Consequences ofFalse-Positive Mammography Screening," *Ann. Fam. Med.*, vol. 11, no. 2, pp. 106–115, 2013.
[68] W. A. Berg, Z. Zhang, D. Lehrer, and R. A. Jong, "Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single Screening MRI to Mammography in Women With Elevated Breast Cancer Risk," *J. Am. Med. Assoc.*, vol. 307, no. 13, pp. 1394–1404, 2012.
[69] P. A. T. Baltzer, M. Benndorf, M. Dietzel, M. Gajda, I. B. Runnebaum, and W. A. Kaiser, "False-positive findings at contrast-enhanced breast MRI: A BI-RADS descriptor study," *Am. J. Roentgenol.*, vol. 194, no. 6, pp. 1658–1663, 2010.
[70] C. Dromain *et al.*, "Evaluation of Tumor Angiogenesis of Breast Carcinoma Using Contrast-Enhanced Digital Mammography," *Am. J. Roentgenol.*, vol. 187, no. 5, pp. W528–W537, 2006.
[71] A. Y. Lee *et al.*, "Inter-reader Variability in the Use of BI-RADS Descriptors for Suspicious Findings on Diagnostic Mammography: A Multi-institution Study of 10 Academic Radiologists," *Acad. Radiol.*, vol. 24, no. 1, pp. 60–66, 2017.
[72] D. L. Weaver *et al.*, "Pathologic Findings from the Breast Cancer Surveillance Consortium Population-Based Outcomes in Women Undergoing Biopsy after Screening Mammography," *Cancer*, vol. 106, no. January 2005, pp. 732–742, 2006.
[73] W. Lu, Z. Li, and J. Chu, "A novel computer-aided diagnosis system for breast MRI based on feature

selection and ensemble learning," *Comput. Biol. Med.*, vol. 83, no. February, pp. 157–165, 2017.
[74] Y. Qiu *et al.*, "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," *J. Xray. Sci. Technol.*, vol. 25, no. 5, pp. 751–763, 2017.
[75] B. K. Patel *et al.*, "Computer-aided diagnosis of contrast-enhanced spectral mammography : A feasibility study," *Eur. J. Radiol.*, vol. 98, no. May 2017, pp. 207–213, 2018.
[76] D. Gur, J. Stalder, and L. a Hardesty, "CAD performance on sequentially ascertained mammographic examinations of masses: an assessment," *Radiology*, vol. 233, pp. 418–423, 2004.
[77] B. Zheng, J. H. Sumkin, M. L. Zuley, D. Lederman, X. Wang, and D. Gur, "Computer-aided detection of breast masses depicted on full-field digital mammograms: a performance assessment.," *Br. J. Radiol.*, vol. 85, no. 1014, pp. e153-61, 2012.
[78] A. Oliver *et al.*, "A review of automatic mass detection and segmentation in mammographic images," *Med. Image Anal.*, vol. 14, no. 2, pp. 87–110, 2010.
[79] M. Tan, F. Aghaei, Y. Wang, and B. Zheng, "Developing a new case based computer-aided detection scheme and an adaptive cueing method to improve performance in detecting mammographic lesions Developing a new case based computer- aided detection scheme and an adaptive cueing method to improve perfor," *Phys. Med. Biol.*, vol. 62, pp. 358–376, 2017.
[80] X. Wang, D. Lederman, J. Tan, X. H. Wang, and B. Zheng, "Computerized Prediction of Risk for Developing Breast Cancer Based on Bilateral Mammographic Breast Tissue Asymmetry," *Med. Eng. Phys.*, vol. 27, no. 3, pp. 934–942, 2011.
[81] G. Danala *et al.*, "Applying Quantitative CT Image Feature Analysis to Predict Response of Ovarian Cancer Patients to Chemotherapy," *Acad. Radiol.*, vol. 24, pp. 1233–1239, 2017.
[82] I. H. Witten, E. Frank, and M. A. Hall, "Data mining: practical machine learning tools and techniques, Third Edition," *Elsevier*, 2011.
[83] Q. Li and K. Doi, "Reduction of bias and variance for evaluation of computer-aided diagnostic schemes," *Med. Phys.*, vol. 33, pp. 868–875, 2006.
[84] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," vol. 23, no. 19, pp. 2507–2517, 2007.
[85] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[86] M. Tan, B. Zheng, J. K. Leader, and D. Gur, "Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development," *IEEE Trans. Med. Imaging*, vol. 35, no. 7, pp. 1719–1728, 2016.
[87] F. Aghaei, M. Tan, A. B. Hollingsworth, and B. Zheng, "Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy," *J. Magn. Reson. Imaging*, vol. 44, no. 5, pp. 1099–1106, 2016.
[88] Y. Wang, F. Aghaei, A. Zarafshani, Y. Qiu, W. Qian, and B. Zheng, "Computer-aided classification of mammographic masses using visually sensitive image features," *J. Xray. Sci. Technol.*, vol. 25, no. 1, pp. 171–186, 2017.
[89] "National Center for Health Statistics. Health, United States, 2017: With special feature on mortality. Hyattsville, MD," 2018.
[90] K. Malhotra, J. Gornbein, and J. L. Saver, "Ischemic Strokes Due to Large-Vessel Occlusions Contribute Disproportionately to Stroke-Related Dependence and Death: A Review.," *Front. Neurol.*, vol. 8, p. 651, 2017.
[91] B. Ray and S. G. Keyrouz, "Management of anticoagulant-related intracranial hemorrhage: an evidence-based review.," *Crit. Care*, vol. 18, no. 3, p. 223, May 2014.
[92] H. P. J. Adams *et al.*, "Guidelines for the early management of patients with ischemic stroke: A scientific statement from the Stroke Council of the American Stroke Association.," *Stroke*, vol. 34, no. 4, pp. 1056–1083, Apr. 2003.
[93] R. G. Nogueira *et al.*, "Thrombectomy 6 to 24 Hours after Stroke with a Mismatch between Deficit and Infarct.," *N. Engl. J. Med.*, vol. 378, no. 1, pp. 11–21, Jan. 2018.
[94] B. R. French, R. S. Boddepalli, and R. Govindarajan, "Acute Ischemic Stroke: Current Status and Future Directions.," *Mo. Med.*, vol. 113, no. 6, pp. 480–486, 2016.
[95] M. Wintermark, R. Sincic, D. Sridhar, and J. D. Chien, "Cerebral perfusion CT: technique and clinical applications.," *J. Neuroradiol. = J. Neuroradiol.*, vol. 35, no. 5, pp. 253–260, Dec. 2008.
[96] T.-Y. Kim, J. Son, and K.-G. Kim, "The recent progress in quantitative medical image analysis for computer aided diagnosis systems.," *Healthc. Inform. Res.*, vol. 17, no. 3, pp. 143–149, Sep. 2011.

[97] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.," *Nat. Commun.*, vol. 5, p. 4006, Jun. 2014.
[98] P. Liu *et al.*, "Applying a radiomics-based strategy to preoperatively predict lymph node metastasis in the resectable pancreatic ductal adenocarcinoma.," *J. Xray. Sci. Technol.*, vol. 28, no. 6, pp. 1113–1121, 2020.
[99] M. Tan, B. Zheng, J. K. Leader, and D. Gur, "Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development.," *IEEE Trans. Med. Imaging*, vol. 35, no. 7, pp. 1719–1728, Jul. 2016.
[100] G. Danala *et al.*, "Applying Quantitative CT Image Feature Analysis to Predict Response of Ovarian Cancer Patients to Chemotherapy," *Acad. Radiol.*, vol. 24, no. 10, pp. 1233–1239, 2017.
[101] G. Danala *et al.*, "Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms," *Ann. Biomed. Eng.*, vol. 46, no. 9, pp. 1419–1431, 2018.
[102] J. Gong, J.-Y. Liu, Y.-J. Jiang, X.-W. Sun, B. Zheng, and S.-D. Nie, "Fusion of quantitative imaging features and serum biomarkers to improve performance of computer-aided diagnosis scheme for lung cancer: A preliminary study.," *Med. Phys.*, vol. 45, no. 12, pp. 5472–5481, Dec. 2018.
[103] F. Aghaei *et al.*, "Implementation of a computer-aided detection tool for quantification of intracranial radiologic markers on brain CT images," in *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, 2017, vol. 10138, pp. 20–27.
[104] J. P. Broderick, O. Adeoye, and J. Elm, "Evolution of the Modified Rankin Scale and Its Use in Future Stroke Trials.," *Stroke*, vol. 48, no. 7, pp. 2007–2012, Jul. 2017.
[105] M. Heidari *et al.*, "Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm.," *Phys. Med. Biol.*, vol. 63, no. 3, p. 35020, Jan. 2018.
[106] X. Chen, A. Zargari, A. B. Hollingsworth, H. Liu, B. Zheng, and Y. Qiu, "Applying a new quantitative image analysis scheme based on global mammographic features to assist diagnosis of breast cancer.," *Comput. Methods Programs Biomed.*, vol. 179, p. 104995, Oct. 2019.
[107] Q. Li and K. Doi, "Reduction of bias and variance for evaluation of computer-aided diagnostic schemes.," *Med. Phys.*, vol. 33, no. 4, pp. 868–875, Apr. 2006.
[108] J. A. Santucci *et al.*, "Radiological Estimation of Intracranial Blood Volume and Occurrence of Hydrocephalus Determines Stress-Induced Hyperglycemia After Aneurysmal Subarachnoid Hemorrhage.," *Transl. Stroke Res.*, Jul. 2018.
[109] Y. Wang, Y. Qiu, T. Thai, K. Moore, H. Liu, and B. Zheng, "A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images.," *Comput. Methods Programs Biomed.*, vol. 144, pp. 97–104, Jun. 2017.
[110] L. Tan, A. Liang, L. Li, W. Liu, H. Kang, and C. Chen, "Automatic prostate segmentation based on fusion between deep network and variational methods.," *J. Xray. Sci. Technol.*, vol. 27, no. 5, pp. 821–837, 2019.
[111] S. D'Souza, "Aneurysmal Subarachnoid Hemorrhage," *J. Neurosurg. Anesthesiol.*, vol. 27, no. 3, pp. 222–240, Jul. 2015.
[112] A. S. Go *et al.*, "Heart disease and stroke statistics--2013 update: a report from the American Heart Association.," *Circulation*, vol. 127, no. 1, pp. e6–e245, Jan. 2013.
[113] A. M. Naidech, J. Drescher, P. Tamul, A. Shaibani, H. H. Batjer, and M. J. Alberts, "Acute physiological derangement is associated with early radiographic cerebral infarction after subarachnoid haemorrhage.," *J. Neurol. Neurosurg. Psychiatry*, vol. 77, no. 12, pp. 1340–1344, Dec. 2006.
[114] B. Ray *et al.*, "Coated-Platelet Trends Predict Short-Term Clinical OutcomeAfter Subarachnoid Hemorrhage.," *Transl. Stroke Res.*, vol. 9, no. 5, pp. 459–470, Oct. 2018.
[115] J. van Gijn, A. Hijdra, E. F. Wijdicks, M. Vermeulen, and H. van Crevel, "Acute hydrocephalus after aneurysmal subarachnoid hemorrhage.," *J. Neurosurg.*, vol. 63, no. 3, pp. 355–362, Sep. 1985.
[116] G. Danala, S. K. R. Maryada, M. Heidari, B. Ray, M. Desai, and B. Zheng, "A new interactive visual-aided decision-making supporting tool to predict severity of acute ischemic stroke," in *Proc.SPIE*, 2020, vol. 11317.
[117] S. Yan, W. Qian, Y. Guan, and B. Zheng, "Improving lung cancer prognosis assessment by incorporating synthetic minority oversampling technique and score fusion method.," *Med. Phys.*, vol. 43, no. 6, pp. 2694–2703, Jun. 2016.
[118] S. Mirniaharikandehei, M. Heidari, G. Danala, S. Lakshmivarahan, and B. Zheng, "Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using CT images," *Comput. Methods Programs Biomed.*, p. 105937, 2021.
[119] H. Baradaran *et al.*, "Evaluating CT Perfusion Deficits in Global Cerebral Edema after Aneurysmal Subarachnoid Hemorrhage.," *AJNR. Am. J. Neuroradiol.*, vol. 36, no. 8, pp. 1431–1435, Aug. 2015.

[120] Y.-X. Ge, J. Li, J.-Q. Zhang, S.-F. Duan, Y.-K. Liu, and S.-D. Hu, "Radiomics analysis of multicenter CT images for discriminating mucinous adenocarcinoma from nomucinous adenocarcinoma in rectal cancer and comparison with conventional CT values.," *J. Xray. Sci. Technol.*, vol. 28, no. 2, pp. 285–297, 2020.
[121] M. E. Buechel *et al.*, "Correlation of imaging and plasma based biomarkers to predict response to bevacizumab in epithelial ovarian cancer (EOC).," *Gynecol. Oncol.*, vol. 161, no. 2, pp. 382–388, May 2021.
[122] T. Wang, J. Gong, H.-H. Duan, L.-J. Wang, X.-D. Ye, and S.-D. Nie, "Correlation between CT based radiomics features and gene expression data in non-small cell lung cancer.," *J. Xray. Sci. Technol.*, vol. 27, no. 5, pp. 773–803, 2019.
[123] R. Sahni and J. Weinberger, "Management of intracerebral hemorrhage.," *Vasc. Health Risk Manag.*, vol. 3, no. 5, pp. 701–709, 2007.
[124] J. L. Ruiz-Sandoval, E. Chiquete, S. Romero-Vargas, J. J. Padilla-Martínez, and S. González-Cornejo, "Grading scale for prediction of outcome in primary intracerebral hemorrhages.," *Stroke*, vol. 38, no. 5, pp. 1641–1644, May 2007.
[125] W. Yang, Y. Feng, Y. Zhang, J. Yan, Y. Fu, and S. Chen, "Volume quantification of acute infratentorial hemorrhage with computed tomography: validation of the formula 1/2ABC and 2/3SH," *PLoS One*, vol. 8, no. 4, pp. e62286–e62286, Apr. 2013.
[126] G. Danala, S. K. R. Maryada, M. Heidari, B. Ray, M. Desai, and B. Zheng, "A new interactive visual-aided decision-making supporting tool to predict severity of acute ischemic stroke," in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2020, vol. 11317, pp. 467–475.
[127] S. Mirniaharikandehei, A. B. Hollingsworth, B. Patel, M. Heidari, H. Liu, and B. Zheng, "Applying a new computer-aided detection scheme generated imaging marker to predict short-term breast cancer risk.," *Phys. Med. Biol.*, vol. 63, no. 10, p. 105005, May 2018.
[128] B. Zheng, Y. Qiu, F. Aghaei, S. Mirniaharikandehei, M. Heidari, and G. Danala, "Developing global image feature analysis models to predict cancer risk and prognosis.," *Vis. Comput. Ind. Biomed. art*, vol. 2, no. 1, p. 17, 2019.