UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

APPLYING NOVEL MACHINE LEARNING TECHNOLOGY TO OPTIMIZE COMPUTER-AIDED DETECTION AND DIAGNOSIS OF MEDICAL IMAGES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MORTEZA HEIDARI

Norman, Oklahoma

2021

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/329479

APPLYING NOVEL MACHINE LEARNING TECHNOLOGY TO OPTIMIZE COMPUTER-AIDED DETECTION AND DIAGNOSIS OF MEDICAL IMAGES

A DISSERTATION APPROVED FOR THE

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Bing Zheng, Chair

Dr. Joseph P. Havlicek

Dr. Hong Liu

Dr. Tomasz Przebinda

© Copyright by MORTEZA HEIDARI 2021

All Rights Reserved.

Acknowledgments

The path toward writing this scientific dissertation would be impossible without the challenges, support, and motivations of kind supervisors, family, people, and friends around me.

First, I would like to express my deep sense of thanks and gratitude to my Ph.D. advisor Professor. Bin Zheng, for his dedication and keen investment in mentoring and encouraging his students toward their passion and finding their path in this long journey. Under his tutelage, I have been motivated to strive for and achieve exceptionally high standards in my research and study. In my study on several research projects, he provided me with outstanding mentorship in the disciplines of initial steps of study design, professional writing protocol, and data analysis. In many areas, both personally and professionally, Prof. Zheng has taught and encouraged me to meet challenges that I had never before thought possible. Time and effort that Professor Zheng invested in developing my scientific curiosity, lead me to accomplish this task.

I owe a deep sense of gratitude to all my committee members, Dr. Joseph P. Havlicek, Dr. Hong Liu, Dr. Tomasz Przebinda, for the personal and professional guidance and suggestions they provided me to complete this task. They were an integral part of me being able to make it to the end. I hope that I can “pay forward” what they’ve done for me by helping other people.

I am grateful to all those with whom I have had the pleasure to work and be friends with, during my Ph.D. study and general life. Seyedehnafiseh Mirniaharikandehei and her husband, Dr. Alireza Abdi, Gopichandh Danala, Dr. Faranak Aghaei, Javad Dowran, Farid Omoumi, Mahammad Abshirini and his wife Prisa Marashizadeh, and Abolfazl Zargari, for their friendship, help, and support.

Nobody has been more important to me in pursuing my study life and accomplishing my Ph.D. than my family members. I am extremely thankful to my mom and my dad, Behjat and

Ahmad; my lovely brothers, Hosein, Ali, and Hasan; my sisters-in-law, Vajihe, Javahar, and Foroozan; and my nieces and nephews whose love, encouragement, and guidance kept me motivated in pursuing my dreams.

To my MOM and my DAD for sacrificing their life to support me!

Table of Contents

List of Tables XI

List of Figures XIV

Abstract XVIII

1 Introduction 1

1.1 Radiology Imaging Modalities 3

1.2 Artificial Intelligence (AI) in the area of medical imaging 5

1.2.1 Machine learning 6

1.2.2 Deep Learning 11

1.3 Organization of the Dissertation 14

2 Research Objective and Hypothesis 16

2.1 Current Status and Challenges of Machine Learning in Medical Imaging Informatics 17

2.2 New Image Features Applying to Computer-aided Detection (CAD) 21

2.2.1 Background 21

2.2.2 Hypothesis and Proposed Approach 24

2.3 Feature optimization/selection and their impacts on CAD performance 25

2.3.1 Background 26

2.3.2 Hypothesis and Proposed Approach 28

2.4 Application of Deep Learning for Optimal Feature Extraction and Classification 30

2.4.1 Background 30

2.4.2 Hypothesis and Proposed Approach 33

3 Development and Assessment of a New Global Mammographic Image Feature Analysis Scheme to Predict Likelihood of Malignant Cases 35

3.1 Introduction 35

3.2 Materials and Methods 37

3.2.1 Image Dataset 37

3.2.2 Background of Image Features 38

3.2.3 Data Preprocessing 42

3.2.4 Image Feature Extraction 43

3.2.5 Classification Phase 47

3.3 Results 50

3.4 Discussion 56

4 Prediction of Breast Cancer Risk Using a Machine Learning Approach Embedded with a Locality Preserving Projection Algorithm 61

4.1 Introduction 61

4.2 Materials & Method 63

4.2.1 Image Dataset 63

4.2.2 A Computer-aided Imaging Processing Scheme 64

4.2.3 Image Feature Computation 66

4.2.4 Machine Learning Generated Imaging Marker 68

4.2.5 Experiments and Performance Evaluation 75

4.3 Results 76

4.4 Discussion 80

5 Applying a Random Projection Algorithm to Optimize Machine Learning Model for Breast Lesion Classification 84

5.1 Introduction 84

5.2 Materials and Methods 86

5.2.1 Image Dataset 86

5.2.2 Initial Image Feature Pool with a High Dimensionality 88

5.2.3 Applying Random Projection Algorithm (RPA) to Generate Optimal Feature Vector 91

5.2.4 Experiment of Feature Combination and Dimensionality Reduction 94

5.2.5 Development and Evaluation of Machine Learning Model 100

5.3 Results 102

5.4 Discussion 107

6 Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms 112

6.1 Introduction 112

6.2 Materials and Methods 114

6.2.1 Dataset 114
6.2.2 Image Preprocessing 115
6.2.3 Transfer Learning 117
6.2.4 Model Training and Testing 119
6.2.5 Performance Assessment 122
6.3 Results 123
6.4 Discussion 128
7 Conclusion and future work 131
7.1 Summary 131
7.1.1 Journal Papers: 134
7.1.2 Conference Proceeding Papers: 136
7.2 Future works 138
References 141

List of Tables

Table 3-1 Case numbers and percentage distribution of patients' age and mammographic density rate by radiologists using BIRADS guidelines. 38

Table 3-2 The computed SSIM, DCT and FFT image Features. 46

Table 3-3 AUC and Accuracy for different sub-group of features on CC view in compare with MLO view. 53

Table 3-4 Computed area under ROC curve using individual group of features on both CC and MLO views. 53

Table 3-5 Four confusion matrices generated using 4 SVMs trained using features computed from both CC and MLO view. 54

Table 3-6 Accuracy, sensitivity, specificity, and odd ratio of using 4 SVMs trained using different features computed from both CC and MLO views. 55

Table 3-7 Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) at five subgroups with increasing values of SVM-generated prediction scores. 55

Table 3-8 Comparison of prediction performance of SVMs trained with and without removal of pectoral muscle area in MLO view images. 56

Table 4-1 Distribution of age and density BIRADS of cases in the dataset. 64

Table 4-2 Description of 44 computed image features in the initial feature pool..... 68

Table 4-3 Accuracy (%) of the whole 44 feature vector for SVM classifiers with different kernel functions. 77

Table 4-4 Confusion matrix of the proposed risk model on 500 cases with Threshold = 0.5. 77

Table 4-5 Odds and Risk Ratio of the proposed KNN-LPP method. 77

Table 4-6 Adjusted ORs and 95 % CIs for five subgroups of cases. 78

Table 5-1 Case numbers and percentage distribution of patients' age and mammographic density rated by radiologists using BIRADS guidelines. 87
Table 5-2 List of computed features on ROI area..... 96
Table 5-3 List of Wavelet-based extracted features..... 98
Table 5-4 List of Geometrical-based extracted features. 98
Table 5-5 Accuracy of the SVM models for case-based classification based on six different categories of the original features. 103
Table 5-6 Summary of the average number of image features used in 5 different SVM models and classification performance (AUC) based on both region and case-based lesion classification. P-value compares the results of each model to the last one (RPA) as the optimal method..... 105
Table 5-7 Five comparison matrices of case-based lesion classification using five different SVM models to classify between benign and malignant cases..... 106
Table 5-8 Summary of the lesion case-based classification accuracy, sensitivity, specificity, and odds ratio of using five SVMs trained using different groups of optimized features. 106
Table 5-9 Summary of the case-based lesion classification for the proposed method (RPA) under different cross-validation (CV) techniques. 107
Table 6-1 The architecture of the new VGG16 model after transfer learning with new layers (19 to 22) 119
Table 6-2 Distribution of cases in three subsets. 120
Table 6-3 Classification report of the proposed method..... 125
Table 6-4 Confusion matrix of four CNN models on X-ray Images. 95% confidence interval (CI) for the accuracy is shown in the last column..... 127

Table 6-5 Comparison accuracy results of the proposed method with the other deep learning methods on COVID-19 diagnosis. 127

List of Figures

Figure 1-1 Machine learning development and application model for medical image classification tasks. For training, the machine learning scheme utilizes a set of input images to identify the image features useful to result in the correct classification of the image like depicting benign or malignant tumor. (b) For predicting, once the scheme has trained in the training images, the optimal model will be applied to the test set to assist radiologists in identifying the tumor type [27] 8

Figure 1-2 A general architecture of CNN. The input is generally an image, and the output could be the image class [42]. 13

Figure 2-1Flowchart of the procedure for short-term breast cancer risk [18]..... 23

Figure 2-2 Overview of an ensemble method [67]. 32

Figure 3-1 Preprocessing phase. a) the original image, b) chest wall removal step, c) denoising black area, and written labels removal. 42

Figure 3-2 Block diagram of the proposed method for SSIM feature extraction. 44

Figure 3-3 Feature extraction phase of the proposed method. 47

Figure 3-4 Correlation coefficient matrices of 34 image features computed from CC (top) and MLO (bottom) view images. 49

Figure 3-5 A summarized block diagram of the proposed scheme for classification of benign and malignant tissues in mammography imaging. 50

Figure 3-6 Illustration of one malignant case (the first row) and one benign case (the second row). The detected masses are circled (Green Color) in the images. 51

Figure 3-7 Comparison of 4 ROC curves generated by four SVMs trained using image features computed from both CC and MLO view images. 51

Figure 4-1 Illustration of the graphic user interface of our computer-aided imaging processing scheme to detect bilateral mammographic image feature asymmetry and predict short-term breast cancer risk. 65

Figure 4-2 An example of showing the intermediate results of image processing steps including (a) computed breast tissue density maps, (b) detected focal density regions, (c) local density (pixel value) fluctuation maps, and (d) image maps generated using Gaussian bandpass filtering. Color bars show volumetric density level of the pixel values. 66

Figure 4-3 Block Diagram of three types of risk model systems, (a) conventional systems for feature selection and classification, (b) deep learning techniques for feature generation and classification, (c) proposed method for feature extraction, regeneration, and classification... 70

Figure 4-4 Accuracy of 44 elements feature vector with KNN classifier system for 10 different K. 79

Figure 4-5 Comparison of two ROC curves generated by the original KNN model using the initial 44 features and the optimal LPP-KNN model using 4 features. 79

Figure 5-1 Example of 4 extracted ROIs with the detected suspicious soft-tissue masses (lesions) in ROI center. a,b) 2 ROIs involving malignant lesions and c,d) 2 ROIs involving benign lesions. 95

Figure 5-2 Example to illustrate lesion segmentation, which include a) the original ROI, b) absolute difference of ROI from low-pass filtered version, c) combination of (a) and (b) which gives the suspicious regions better contrast to the background, d) output of of morphological filtering, e) blob with the largest size is selected (a binary version of the lesion), and f) finally segmented lesion area. It is output of mapping (e) to (a). 96

Figure 5-3 Wavelet based feature extraction. Wavelet decomposition is applied three times to make the images compress as possible. Then PCA is adopted as another way of data compression. 98
Figure 5-4 Illustration of the overall classification flow of the CAD scheme developed and tested in this study. 100
Figure 5-5 A malignant case annotated by radiologists in both CC and MLO views. The annotated mass is squared in each view. 103
Figure 5-6 A trend of the case-based classification AUC values generated by the SVM models trained using different number of features (NF) generated by the RPA. 104
Figure 5-7 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. 105
Figure 6-1 Example of chest X-ray images in three classes (top – normal, middle – community-acquired non-COVID-19 pneumonia, and bottom – COVID-19 infected pneumonia case). The figure also shows (a) the original Images, (b) the binary images after threshold, (c) images after selecting the biggest segmented region, (d) images after applying morphological filtering and (e) the original image after removing the majority part of diaphragm region (I_p). 116
Figure 6-2 A flow diagram to illustrate image pre-processing steps to generate input of a CNN model, where (I) is the original Image in the dataset. (I_p) is the diaphragm removed image. (I_eq) is an image after applying histogram equalization on (I_p), and (I_b) is an image after applying bilateral filtering on (I_p). Three images (I_p), (I_b), and (I_eq) are fed into three channels of the CNN model to simulate the RGB image. 117
Figure 6-3 Illustration of the architecture of VGG16 based CNN model. 119

Figure 6-4 schematic representing training and validation phase of the proposed scheme. .. 122
Figure 6-5 (a-c) Left column show three sets of performance curves applying to the training and validation subsets for 3 experiments in 200 training epochs, respectively. The horizontal axis shows the number of epochs, and the vertical axis shows the accuracy. The right column show three confusion matrices of the corresponding testing results are shown on the right. (d) A combined confusion matrix of applying three trained models to three independent testing data subsets with total 2,544 cases. 125

Abstract

The purpose of developing Computer-Aided Detection (CAD) schemes is to assist physicians (i.e., radiologists) in interpreting medical imaging findings and reducing inter-reader variability more accurately. In developing CAD schemes, Machine Learning (ML) plays an essential role because it is widely used to identify effective image features from complex datasets and optimally integrate them with the classifiers, which aims to assist the clinicians to more accurately detect early disease, classify disease types and predict disease treatment outcome. In my dissertation, in different studies, I assess the feasibility of developing several novel CAD systems in the area of medical imaging for different purposes. The first study aims to develop and evaluate a new computer-aided diagnosis (CADx) scheme based on analysis of global mammographic image features to predict the likelihood of cases being malignant. CADx scheme is applied to pre-process mammograms, generate two image maps in the frequency domain using discrete cosine transform and fast Fourier transform, compute bilateral image feature differences from left and right breasts, and apply a support vector machine (SVM) method to predict the likelihood of the case being malignant. This study demonstrates the feasibility of developing a new global image feature analysis based CADx scheme of mammograms with high performance. This new CADx approach is more efficient in development and potentially more robust in future applications by avoiding difficulty and possible errors in breast lesion segmentation. In the second study, to automatically identify a set of effective mammographic image features and build an optimal breast cancer risk stratification model, I investigate advantages of applying a machine learning approach embedded with a locally preserving projection (LPP) based feature combination and regeneration algorithm to predict short-term breast cancer risk. To this purpose, a computer-aided image processing scheme is applied to segment fibro-glandular tissue depicted on mammograms and initially compute 44 features related to the bilateral asymmetry of

mammographic tissue density distribution between left and right breasts. Next, an embedded LLP algorithm optimizes the feature space and regenerates a new operational vector with 4 features using a maximal variance approach. This study demonstrates that applying the LPP algorithm effectively reduces feature dimensionality, and yields higher and potentially more robust performance in predicting short-term breast cancer risk. In the third study, to more precisely classify malignant lesions, I investigate the feasibility of applying a random projection algorithm to build an optimal feature vector from the initially CAD-generated large feature pool and improve the performance of the machine learning model. In this process, a CAD scheme is first applied to segment mass regions and initially compute 181 features. An SVM model embedded with the feature dimensionality reduction method is then built to predict the likelihood of lesions being malignant. This study demonstrates that the random project algorithm is a promising method to generate optimal feature vectors to improve the performance of machine learning models of medical images. The last study aims to develop and test a new CAD scheme of chest X-ray images to detect coronavirus (COVID-19) infected pneumonia. To this purpose, the CAD scheme first applies two image preprocessing steps to remove the majority of diaphragm regions, process the original image using a histogram equalization algorithm, and a bilateral low-pass filter. Then, the original image and two filtered images are used to form a pseudo color image. This image is fed into three input channels of a transfer learning-based convolutional neural network (CNN) model to classify chest X-ray images into 3 classes of COVID-19 infected pneumonia, other community-acquired no-COVID-19 infected pneumonia, and normal (non-pneumonia) cases. This study demonstrates that adding two image preprocessing steps and generating a pseudo color image plays an essential role in developing a deep learning CAD scheme of chest X-ray images to improve accuracy in detecting COVID-19 infected pneumonia.

In summary, I developed and presented several image pre-processing algorithms, feature extraction methods, and data optimization techniques to present innovative approaches for

quantitative imaging markers based on machine learning systems in all these studies. The studies' simulation and results show the discriminative performance of the proposed CAD schemes on different application fields helpful to assist radiologists on their assessments in diagnosing disease and improve their overall performance.

1 Introduction

Medical image interpretation is the principal undertaking of radiologists to detect and diagnose diseases at their early stages and treat them before late. While the quality of the image and image interpretation are two main factors, image interpretation by radiologists is limited by distractions like visual search pattern, fatigue, noise, and even analyzing a large number of images in the dataset, which makes it nearly impossible to analyze them by radiologists. So, the demand for a machine or computational method as a "second-reader" to analyze medical images faster and assist radiologists became obvious. Hence, the early disease diagnosis and prognosis assessment have attracted great clinical research interest over the past decades, facilitating patients' subsequent clinical management. The importance of diagnosing the presence of diseases for patients at an early stage has led many research teams in the biomedical imaging informatics field to study different techniques to develop new computer-aided detection and diagnosis schemes (CAD) of medical images and improve their performance in the field. The purpose of developing computer-aided detection (CADe) schemes is to assist physicians (i.e., radiologists) to more accurately interpret medical imaging findings and reduce inter-reader variability [1]. Although previous observer performance studies reported that using CADe might help radiologists detect more diseases that may be previously missed or overlooked by radiologists (i.e., [2]), the clinical data analysis studies showed that using CADe increased false-positive recalls and reduced radiologists' performance measured by areas under the receiver operating characteristics curves (i.e., [3]). Thus, the specificity of current medical imaging schemes remains lower in clinical practice. The higher false-positive recall rates add anxiety with potentially long-term psychosocial consequences [4] and physical harm to many people who participate in multiple screening and unnecessary biopsies [5]. The high false-positive recall rates also associate with a high economic burden on the healthcare system [6]. Thus, in order to further improve the efficacy of screening systems, developing the computer-aided

diagnosis (CADx) schemes that aim to assist radiologists in their decision-making to better assess the risk of the detected suspicious area as the area of infection and reduce the unnecessary biopsies of clean lesions have been attracted broad research interest for the last two decades [7].

Despite excellent research effort, CADx schemes have not been broadly accepted and utilized in clinical practice. It still faces multiple technical challenges to improve CADx performance and robustness. For instance, previous strategies, which are popular as radiomics methods, typically include three steps, each with a wide range of challenges, namely (1) collecting data, implementing image processing algorithms, and segmenting the region of interest depicting on medical images, (2) computing hand-crafted image features from the segmented regions, and (3) train multi-feature fusion-based machine learning classifiers for training and validation to predict a specific clinical endpoint like tumor segmentation, response to treatment, or risk prediction [8].

Radiomics is a term presented by Lamb in [9] to represent image feature extraction from radiological data. The goal of radiomics feature extraction is to connect the human vision system, features, and outcomes in a meaningful way to improve precision and provide trustable results. In radiomics based methods, we actually derive many features from medical images to mimic the characteristics of the diseases that are hard to get by naked eyes, which would be considered complementary information. However, due to different medical image challenges, accurate lesion segmentation is often tricky and unreliable, which can substantially affect the performance and robustness of CADx schemes [10]. Despite promising outcomes that recent efforts have led to significant developments on any of these aspects, radiomics methods have yet to be efficiently applicable in clinical practices since the complexity of these methods' workflow has ended in a lack of standardization. The feature extraction techniques, strategies, and the number of features are all involving constantly without any established standard behind them in these methods. Hence, numerous obstacles like dimensional reduction, object segmentation, and redundancy interpretations remained open problems to the researchers [11].

To overcome this difficulty, researchers recently investigated and applied deep learning techniques to develop CADx schemes without lesion segmentation and hand-crafted feature computation as two main challenges [12, 13]. Although deep learning approaches can avoid difficulty in lesion segmentation and manually defining image features, they require a large and diverse image dataset of one of the radiology imaging modalities to train the scheme to minimize the risk of overfitting, and validate its performance, which is another tedious task hard to get. The reason is that in the area of medical imaging, radiology imaging modalities are not that much large and diverse in comparison to the other area. Considering the fact that based on the specific purpose that radiologists have, they use a particular imaging modality to screen inside the body.

1.1 Radiology Imaging Modalities

Radiology imaging modalities are practical to screen internal body structure for different purposes like diagnosing disease, analyze the body's response to a particular treatment, cancer staging, and detection, bone injury, brain injury discovery, etc. The most well-known radiology imaging modalities are plain x-ray, magnetic resonance imaging (MRI), computed tomography (CT), ultrasound imaging, and positron emission tomography (PET).

In X-ray imaging systems, X-ray beams, which are high-energy electromagnetic radiations, pass through the body, and in their route, depending on the density of the material they pass through, they will be absorbed in different amounts. The passed-through part of the beams will reach the detector and produce a 2D X-ray image.

CT is a combination of X-ray images projected from different angles. With computer processing algorithms, all these images are putting together to create 3D images from structures inside the body. Thus, it is clear that CT images provide information with more details from inside the body compared to 2D X-ray. Muscles and bone structures, internal bleeding, and surgical procedures are all possible to screen with CT imaging. CT imaging is a popular imaging

modality in the medical imaging field for diagnosing diseases, which has also resulted in the enormous increase in applying CT imaging to screen, detect, and diagnose lung cancer and other lung diseases (i.e., COPD). The challenge and difficulty of detecting and classifying subtle lung nodules by radiologists have promoted broad research interests in developing computer-aided detection and/or diagnosis (CAD) schemes of thoracic CT images [14].

MRI is a medical imaging method that provides an image from inside the body's organs and tissues based on a magnetic field, magnetic field gradient, and radio waves. This process happens in large and tube-shaped machines. MRI typical is beneficial to diagnose tissues and skeletal system.

Ultrasound uses high-frequency sound waves with frequencies higher than the upper limit of human hearing to produce images from organs inside the body. In this technique, the ultrasound machine connected to a transducer sends high-frequency sound waves to the body, then the organs inside the body reflect the waves, the device captures them and with a computer algorithm generates real-time images of the organs.

PET scan is a type of imaging technique that uses radioactive substance injected through a vein as a tracer to track the disease inside the body and figure out how organs and tissues are working. The radioactive substance will be absorbed into the target area, which is essential for the radiologist to see details clearly. After the injection, the patient will go to the PET scanner, and the PET detects signals coming from the tracer. With computer algorithms, the signals will be processed as 3D images.

Each of these imaging modalities has its negative sides while provides valuable information for specific diseases. For instance, x-ray images are easy to operate, while x-ray has the risk of radiation exposure to patients. In contrast, there is no radiation exposure to the patients under ultrasound imaging, but low penetration through bone makes it tricky to interpret. If the radiologist aims to track anatomical variations in the tissues, MRI and CT imaging are

appropriate options. MRI presents a better resolution for soft tissues, but it has a time-consuming scanning process, so for patients in need of urgent care, it may not be advisable. If the purpose is to capture biochemical variations, PET imaging is the right preference [15]. Therefore, it is rational that radiologists, based on the purpose and the available resources, utilize each of which for their prognoses.

CAD schemes have been developed to help address many challenging tasks based on any of these imaging modalities. For instance, pulmonary embolism segmentation with computed tomographic (CT) angiography [16], polyp diagnosis with virtual CT or colonoscopy in the setting of colon tumor [17], breast cancer detection and diagnosis with mammography [18], brain tumor segmentation with magnetic resonance (MR) imaging [19], tumor treatment efficacy assessment of mouse model using ultrasound images [20], brain abnormality detection from PET images [21], or early diagnosis of Alzheimer's disease (AD) based on machine learning techniques on PET and MRI based features [22], and diagnosis of the cognitive state of the brain with functional MR imaging to diagnose neurologic disorders [23].

In developing CAD schemes, Artificial Intelligence (AI) plays a critical role because it is widely used to identify effective image features from complex datasets and optimally integrate them with the classifiers, which aims to assist the clinicians to more precisely detect early disease, classify disease types and predict disease treatment consequence. For example, in [24], with the application of machine learning techniques, the accuracy of cancer prediction outcomes has significantly improved by 15%–20% during the last years.

1.2 Artificial Intelligence (AI) in the area of medical imaging

The development of computers and the advance in imaging have made the emergence of artificial intelligence (AI) impressive in the area of medical imaging. The artificial neural network (ANN) is a resolution to machine learning (ML) and deep learning (DL), which could be basically a combination of structural layers of connected nodes [25]. The input to an ANN

could be radiomics features that have been derived from medical images, or it could be the whole medical image if we take advantage of convolutional neural networks (CNN).

The capability of AI in the area of medical imaging includes a wide range of applications on lesion detection/classification, medical imaging analysis/segmentation, image retrieval, image reconstruction, and radiomics feature extraction to extract abstract features that the vision system is not capable of tracking them. To do so, ML conducts a more specific interpretation to consolidate more data capable of representing a higher level of generality. For the plan, an ANN will be configured based on the nodes of an appropriate number of layers. Nodes in the deeper layers get information from previous nodes, and combine them to get to the output as the last layer. The purpose, which is maximizing the accurate answer based on ground truth, is achieved by adjusting the nodes' weights with optimization algorithms under different iterations. The error associate with the precise answer is calculated under a pre-defined loss index, and it will be optimized during the training process.

1.2.1 Machine learning

As a division of Artificial Intelligence (AI), machine learning is proper to address and determine the problem of learning patterns from data samples to the general concept of inference [26]. Every learning process consists of two main phases: (i) estimation of unknown dependencies in a scheme from a provided dataset and (ii) management of estimated dependencies to predict new outputs of the scheme. During the implementation of an ML method, data samples are provided to develop the essential components. Every instance is described with several features, and every feature can hold a wide range of values. Figure 1-1 shows one broadly-accepted illustration of a machine learning scheme: If a machine learning algorithm is implemented for a set of data (and in a supervised manner to some knowledge about these data, for example, benign or malignant tumors), then the algorithm scheme can learn from the training data and take advantage of what it has learned to make a

prediction/classification on the future data [27]. We commonly train the designed ML on a generally large and diverse dataset with a loss function. Then, a second independent dataset is prepared to validate the performance of ML if it is trained well or not. Although the loss function estimates the error between the outputs and the ground-truth of training data, under a large and diverse training dataset, this error will be generalized to the new data, and the ML would work well for any further test data [28]. Typically, in the training and validation process, the training phase has high performance (from the loss function point of view) and better than the test phase. When this difference is small, results show the ML is trained well and not suffering from overtraining.

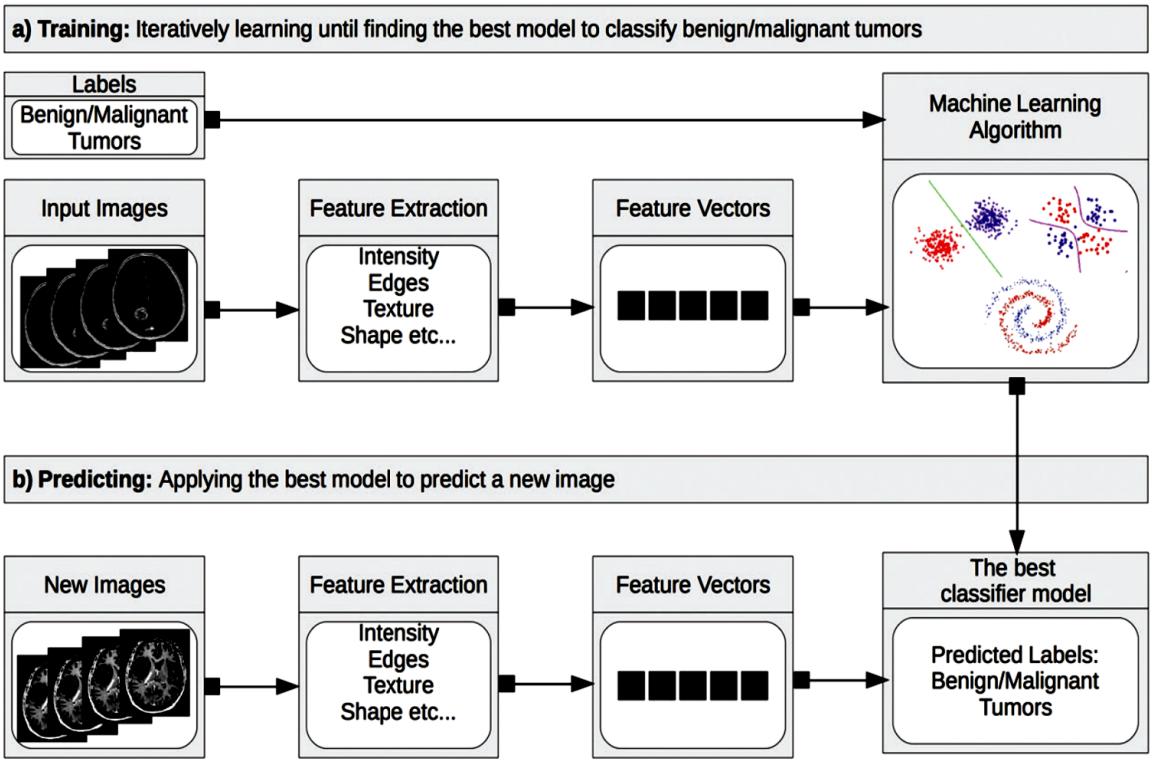


Figure 1-1 Machine learning development and application model for medical image classification tasks. For training, the machine learning scheme utilizes a set of input images to identify the image features useful to result in the correct classification of the image like depicting benign or malignant tumor. (b) For predicting, once the scheme has trained in the training images, the optimal model will be applied to the test set to assist radiologists in identifying the tumor type [27].

However, robust trained ML models are hard to get because of different limitations identified in the area of medical imaging—first, a perfect image segmentation that is like a bottleneck to any excellent radiomics analyses. Second, the large number of handcraft features, since for small datasets, the larger size of feature set reduces performance reliability. In other words, for each extracted feature, we should have at least 5 cases per class to do a reliable learning process. Hence, since providing large datasets in the medical imaging area is a challenge itself, we need to figure out how to select the feature-set from thousands radiomics features optimally [11]. Consequently, the determination and optimization of ML for model training and validation remain an open problem. Meanwhile, different studies have shown that while there is not a single best classifier with optimal parameters across all datasets, choosing

an appropriate classifier, a proper training style, and an appropriate selection of features (feature engineering) can provide more trustable results [29, 30].

1.2.1.1 Classifiers and training styles

Machine learning algorithms can be categorized based on training styles: supervised, unsupervised, semi-supervised, and reinforcement learning. When it is about supervised learning, we have a dataset along with the labels. Labels are used as answer keys to the algorithm during the training process to evaluate the training data's accuracy and optimize the model as far as possible. In unsupervised learning, on the other hand, training happens based on an unlabeled dataset. During the training process, the algorithm attempts to make sense of the dataset by independently deriving features and patterns. Semi-supervised learning is like a middle-ground process; it uses a small amount of labeled data and a large amount of unlabeled data for the training process. Reinforcement learning does the training step based on a rewarding system to provide feedback for the best purpose in particular situations. Hence, the optimal strategies are gained based on an optimal way to accomplish one specific goal. More specifically, reinforcement learning is based on interaction with the environment.

It is possible to employ support vector machine, decision tree, linear regression, logistic regression, naive Bayes, k-nearest neighbor, random forest, AdaBoost, and neural network methods for supervised learning algorithms. In comparison, unsupervised learning algorithms are habitually based on K-means, mean shift, hierarchical clustering, affinity propagation, DBSCAN (density-based spatial clustering of applications with noise), Markov random fields, Gaussian mixture modeling, ISODATA (iterative self-organizing data), and fuzzy C-means schemes. General adversarial network (GAN) is a popular scheme for the semi-supervised learning process, and finally, for the reinforcement learning category, we can mainly consider deep learning approaches [31-33]. Overall, based on the amount of data assembled and, most

importantly, their corresponding labels, we should select an appropriate classifier and the proper training style to achieve optimal results.

1.2.1.2 Feature engineering

In the process of feature extraction step, some data-related issues refer to the quality of the data, and the preprocessing steps are required to make them more suitable for ML. Data quality issues include the appearance of noise, outliers, missing or duplicate data, saturated data, and data that is biased-unrepresentative. Several distinct techniques and strategies exist in the medical imaging informatics field, referring to feature preprocessing that focuses on transforming the data to fit a specific ML method better. Among these ways, some of the most critical approaches include feature extraction, dimensionality reduction, and feature selection.

For the feature extractions, there are three foremost types of features for image analysis. The first one is the geometric features, such as geometric shape features, radius features, profile features, the boundary and circularity information, major and minor axes and their ratio, etc. The second type of features is textural features, such as run-length features, local binary patterns (LBP), cooccurrence features, vector quantization generating texture descriptor, the histogram of oriented gradients (HOG) features, and wavelets. The third category of features is intensity-based ones. Like gradient magnitude features, edge-gradient features, CT value histogram (CVH), and intensity distributions [34-36].

Although it is feasible to compute many features from an image, driving too many features can push the scheme to overfit rather than learning the actual support of a determination. The process of choosing the subset of features that are utilized to obtain the best prognostications is known as feature selection [37]. There are many advantages regarding the dimensionality reduction and feature selection when we extract a large number of features. ML algorithms accomplish better when the dimensionality is lower. Additionally, the reduction of dimensionality can exclude irrelevant features, reduce noise, and can generate more robust

learning paradigms due to the involvement of fewer features. In general, if the dimensionality reduction happens by choosing new features that are a subset of the former ones, it is identified as feature selection. Three main techniques are available for feature selection, namely embedded, filter, and wrapper strategies.

Wrapper techniques assess the utility of feature subsets utilizing the results of a defined classifier. These methods recognize the possible interactions between variables. Then, an exploration procedure within the possible feature subsets space is done. As the number of subsets expands exponentially with the number of features, a heuristic or a sequential selection algorithm is used for search purposes. The two principal disadvantages of these techniques are the increasing overfitting risk when the number of observations is inadequate and the meaningful computation time when the number of variables is considerable [38].

Embedded methods diminish computational time compared to wrapper methods by incorporating feature selection in a classifier's training process. They are very sensitive to the learning algorithm employed to set feature subsets. Support Vector Machine (SVM) based approaches or decision-tree-based algorithms are some examples appropriate for embedded methods [39].

Filter methods do not depend on any classifier but can be acknowledged pre-processing steps based on particular criteria to evaluate features' relevance. One of these approaches' main disadvantages is that they ignore the interaction between features and may not remove redundant features. The most recommended techniques are univariate. This means that each component is considered in the evaluation process separately, for instance, on mutual information [40].

1.2.2 Deep Learning

A new class of AI has proposed and developed in 2007 is named deep learning algorithm. They became famous in 2012 while one of the approaches based on CNN won the best computer

vision competition, ImageNet [41]. They are different from classical ML that is based on extracted features as their input. DL inputs are images and use pixel values in the images directly rather than feature calculation; thus, they don't require segmentation, feature extraction, and feature optimization techniques still known as challenging open problems in the area of medical imaging. For instance, for complicated lesions in a complex background, a perfect automatic segmentation is always challenging, or calculating relevant handcraft features for a given task is hard to get. While DL can avoid relevant errors caused by these steps; hence, DL's performance is generally higher than classical ML methods. In DL-based algorithms, all required steps to the given task are done in multiple nonlinear hidden layers. In other words, DL bypasses all segmentation, feature calculation, and optimization steps to ascertain high-level features for the relevant task. Gradually under different iterations, a distinguished representation of the task is determined by DL. DL is named end-to-end algorithm, since it does the whole process from input images to the final classification fully automatic.

Convolutional Neural Network (CNN) is a subclass of deep learning typically utilized to analyze images. A general architecture of CNN is illustrated in Figure 1-2. As explained earlier, the inputs are images, and the outputs could be either class categories or a segmented part of the input images. The layers of CNN are connected either with local shift-invariant pooling and convolutional layers. The neurons in every layer are connected to the local neurons in the previous layer. So, the forward propagation data are like a shift-invariant convolutional operation [42]. Convolutional layers in CNN act like feature extraction filters. Each CNN takes advantage of multiple convolutional layers produced by kernels. Feature maps generated by these layers are passed through activation functions to reduce the change of saturation. Then with pooling layers, the feature maps would be down-sampled to reduce the computational requirements. A sequence of multiple of these layers will make the core of CNN, then by fully connected layer, the results provided by all convolution/pooling layers are put to getter to

deliver the classification purpose [25]. The most famous CNNs to date are VGG16/19 [43], ResNet [44], Xception [45] , Inception [46], NasNetLarge [47].

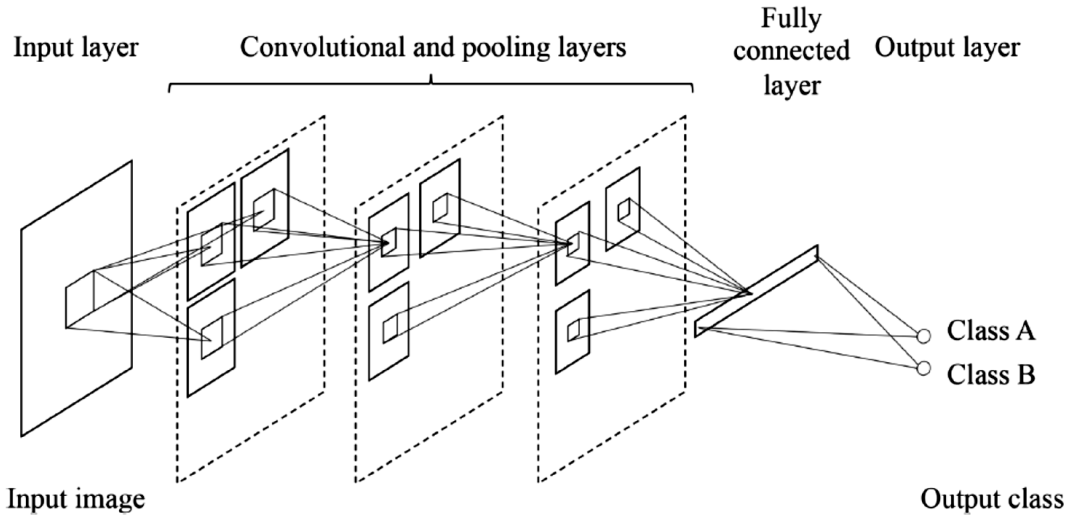


Figure 1-2 A general architecture of CNN. The input is generally an image, and the output could be the image class [42].

For the training process, to have a proper convergence for CNN models, we need to determine hyperparameters like learning rate, optimization algorithm, dropout rate, batch size, and regularization parameters properly. Moreover, the introduction of backpropagation was helpful as an efficient way to optimize the training loss concerning the model parameters. However, the size of the dataset is a challenge to train DL models. In other words, since the number of trainable parameters in DL systems is considerably large (VGG16, 138 million, Resnet152, 60 million, NASNetLarge 89 million), so for small datasets, there is a risk of overfitting, and the DL will memorize the training dataset instead of learning the actual pattern for the task.

One straightforward way to deal with overfitting to the dataset for CNNs is by collecting more data. However, since datasets are not that much large in radiology, training CNNs from scratch is often not possible for CADe/x for medical imaging purposes. Hence transferring genetic features (i.e. shape, edge, and general characteristics) of CNN already trained on natural

scenes, utilizing them as initial values of the features of the CNN as a classifier focused on the medical imaging task, has shown breakthrough in recent years [48]. This technique has two main advantages. The first one is overcoming the shortage of large datasets in the area of medical imaging, and the second one is the computational cost requirement of the CNN for the training process. Another practical way is data augmentation [49] techniques like shearing, shifting, cropping, additive noise, flipping, rotation, etc., proposed for DL-based models to increase the training dataset's size.

In the medical imaging area, CNN has been developed for different tasks, including colonic polyps detection [50], or long decease diagnosis [51] using CT images, breast tomosynthesis analysis using x-ray [52], ultrasound [53], MRI [54], etc.

1.3 Organization of the Dissertation

Previous studies have been conducted to improve overall CAD scheme's performances by optimizing their every stage performance discussed in the previous subsections. In the next chapter, I will review and discuss the state-of-the-art research approaches on feature extraction and feature selection techniques and performance optimization of ML/DL applications to improve the CAD performance schemes presented in the recent medical imaging informatics studies field scheme. Besides, I have conducted many research pieces in the last several years at the University of Oklahoma CAD lab on the related topics. I will present them as my solutions to the challenges.

This dissertation reports four applications of developing CAD schemes using medical imaging modality to segment disease regions (ROIs), extract and optimize features, and apply different machine learning algorithms to perform the classification tasks (Chapter. 3,4,5,6). Specifically, in Chapter 2, a brief of previous studies will be reported on each challenge, then I will propose some solutions to some of the open problems as well. After that, in the next chapters, I will explain each of my recommendations in detail.

In Chapter 3, a new quantitative image processing and feature extraction method will be developed and assessed to classify benign and malignant lesions in mammograms. Chapter 4 will mainly focus on the feature optimization technique and propose a new quantitative global breast feature analysis scheme developed to predict tumors in the next screening of breast images. In Chapter 5, a computer-aided detection tool is generated and optimized based on a random projection algorithm to classify breast lesions using mammogram images. In Chapter 6, a computer-aided detection scheme based on a deep neural network is developed to automatically identify the presence of Covid-19 in chest x-ray images. Last, in Chapter 7, a summary of these new CAD schemes and their future application potentials are discussed, which generates this dissertation's conclusion, including this candidate's future work.

2 Research Objective and Hypothesis

In short, in the previous chapter, I presented that the purpose of developing CAD schemes is to assist physicians (i.e., radiologists) in interpreting medical imaging findings and reducing inter-reader variability more accurately. ML plays an essential role in developing CAD schemes because it is widely used to identify compelling image features from complex datasets and optimally integrate them with the classifiers, which aims to assist the clinicians in more precisely detecting early disease, classify disease types and predict disease treatment outcome.

In the process of designing ML applications on medical images, it is critical to recognize the central application around which ML is to be designed. Then we need to optimize the whole system for the purpose [55]. There are many applications possible to develop based on ML algorithms. For instance, it is possible to develop ML for image acquisition to reduce the processing time by recognizing the target lesion and suggest modification and improvements during the acquisition [56]. For automated detection applications, ML is performed to detect fractures, and organ laceration [57]. ML has also been developed to reduce pulmonary nodules' false-positive results [58], or prostate cancer detection based on MRI [59]. Bone age analyses are another category of applications that can be developed based on ML algorithms to help replace manual workflow [60].

Another potential use of machine learning is for improving the interpretation of findings for breast cancer diagnosis [61], Pulmonary image analysis for different determinations of classification, detection, and segmentation [62], breast lesion detection, segmentation, staging, prediction in digital mammograms and MR imaging [63, 64], lung cancer detection, classification using PET/CT/Microscopic images [65], lesion image synthesis using X-ray/CT images [66], automatic segmentation of multiple organs using 3D CT images [67], skin disease

abnormality analysis based on OCT imaging [68], echocardiography view classification using ultrasound images [69], etc.

It is even possible to take advantage of machine learning to extract related lost information of CT images from MRI images [70]. Medical image registration [71], quantitative assessment of structures [72], and automatic segmentation for brain MRI images [73]. They are all the other applications possible to develop based on machine learning techniques and algorithms. Each of these applications has its barriers and challenges.

To present a robust and reliable ML algorithm for any of these applications, optimizing the performance related to each of the three main stages is worthy of consideration. These stages are (1) image processing and segmentation for a proper feature extraction domain and identify practical medical imaging features, (2) feature optimization and data reduction techniques, and (3) optimizing parameters of ML/DL-based models to improve CAD performance. To present practical algorithms for optimizing each of these steps and make a positive impact on the outputs, we have to initially define an overflow of the application, and the purpose of the application, then discover what would be the difficulties and challenges that mainly overpower the efficacy of the CAD scheme, and finally develop an optimal method or strategy to solve the problem. In the following section, I will present the current challenges that ML-based techniques are faced with for each of these steps. Then I will discuss some of the challenges based on some already available research and studies and present their solutions. Finally, I will provide my proposed hypothesis and solution briefly and then discuss them in more detail in the next chapters.

2.1 Current Status and Challenges of Machine Learning in Medical Imaging Informatics

The development of CAD schemes is based on machine learning algorithms as their core. When the core is a conventional machine learning method, it is necessary to present practical

image processing algorithms to remove redundant information, noise, and artifacts and improve the disease pattern's quality compared to the background. In this way, it is much easier to distinguish different types of structures, and textural information in the images, and extract powerful features that are capable of modeling these differences much better. Then, the next challenge for developers is designing feature extraction techniques based on the domain knowledge to present an appropriate CAD scheme. For both of these two steps, the imaging modality initially creates several obstacles. For instance, image processing objections for 3D CT images are much different from 2D X-ray images, so the solution that works well on one of them may or may not help the other modality. It means each CAD scheme based on each of the imaging modalities has many barriers for the ML-based strategy to work well. Some of the challenges are general, but some are specifically for the proposed method that has to investigate well.

Each imaging modality provides images from the body or the lesion area with varying qualities. Besides, sometimes for the same imaging modality, various imaging machines in different sites may give different image quality because of setups and their imaging performance. Furthermore, structures and patterns of lesions in different imaging modalities would be distinct. For instance, in CT images, pixel values for the body's tissues like bone, fat, and organs have a particular range, while in x-ray images, it is not the same, or the level of noise on X-ray images is higher than CT images, so the lesion pattern may not be that much clear. Also, in 3D CT images, it is possible to find the exact location of the lesion in the body, or examine the volume of the lesions and extract 3D-related features for the ML algorithm, or it is even possible to analyze inter-frame features and characteristics, while it is not possible in 2D x-ray images to go for these categories of features. Thus, the imaging modalities' limitations are different, and it is nearly impossible to propose a unique general image processing or feature extraction method applicable to all these imaging modalities. Even on a particular imaging modality, ML-based schemes are developed to detect, classify, or diagnose different patterns,

lesions, and abnormalities that have not the same characteristics. In other words, based on the purpose, the image processing algorithm would be different. It signifies that we have to review each imaging modality's capability, negative and positive points, then present a practical image processing solution to that modality's challenges to improve the proposed method's robustness in the clinical practices experiences.

Furthermore, in the feature extraction step, there are many challenges to face with. For instance, since the pixel range of different imaging modalities is different, initial setups for PDF and textural-related feature extraction should be different. After making the matrix of features, since they are handcraft features and many parameters are not evaluated well, feature engineering and evaluation are worthy of consideration. In fact, because of possible artifacts, noise, unrelated patterns, and many other unexpected parameters during feature extraction, some of the features may not have related information. They may act like noise with a negative impact on the results, so it would be necessary to find and remove them from the initial feature set. There is a wide range of feature optimization techniques and algorithms appropriate for analyzing the feature sets, all with negative and positive sides. Since the logic behind their operational algorithm is not the same, and they are sometimes dependent on the input feature vector or the selected ML scheme, or even the training algorithm, so it is another open problem for developers to distinguish the optimal data optimization technique to analyze the initial feature vector and optimize it for the CAD scheme.

Another category of challenges comes with choosing a suitable machine learning algorithm for the CAD scheme. For example, if the dataset is not initially annotated, we have to pick unsupervised learning algorithms like C-means and K-means. Similarly, based on the amount of annotated data, different semi-supervised and supervised learning methods are available, which going for appropriate classifier and even learning algorithm is still an open problem. Also, even for a particular ML method developing an optimal strategy for hyperparameter tuning during the training step is another challenge.

While DL-based CAD schemes are presented to bypass the image processing and feature extraction steps and their challenges, they bring a new division of objections for developers. For instance, in DL, the chance of underfitting because of not enough data, overfitting because of learning too many details and noise in the data, overtraining to one of the classes because of imbalance dataset, and memorizing the training dataset instead of the actual pattern because of the large size of trainable parameters in DL algorithms is high in comparison with classical ML systems. Also, the process of tuning and achieving the optimal hyperparameters is sometimes unapproachable in DL-based techniques since the number of hyperparameters is much more compared with ML algorithms. Besides, defining a perfect loss function to converge the DL algorithm to the optimal results is another difficult task that needs precise analyses to tackle. Furthermore, DL-based CAD schemes' training process is time-consuming and hard to get with conventional computers, so powerful GPU hardware is necessary for them.

Moreover, in the area of medical imaging, datasets are a big challenge to DL-based algorithms. In medical imaging informatics, because of the limitations like patient privacy and data collection process, datasets are not that large and diverse like the other area of researches. Also, since the number of patients with the disease is usually much less than normal cases, imbalanced data is a vital objection for DL-based algorithms. Imbalance data can cause overtraining to the larger class, or when the number of cases in the smaller class is much less, it can cause undertraining, which both means a system that is not trained well. Hence, directly utilizing DL-based algorithms to the dataset to provide CAD schemes may not present trustable outcomes. Figuring out how to deal with each of these obstacles and challenges to overcome the adverse effects and develop high-performance CAD schemes acceptable by radiologists opens many research areas to the developers.

In my dissertation, I am going to investigate some of these challenges for different purposes and applications in more detail, then propose my own hypothesis for an approach and analyze how my solution can overcome the challenge on one of the image modalities.

2.2 New Image Features Applying to Computer-aided Detection (CAD)

In the previous part, I explained that proper feature extraction methods play a critical role in presenting a robust CAD scheme. Still, a wide range of obstacles should be tackled to accomplish practical and effective features. In this part, first, I will present some schemes that their main contribution is on feature extraction techniques. Then, in more detail, I will describe two different studies that are basically based on feature extraction ways on mammography, and ultrasound images as two separate medical imaging modalities. One of them mainly offers a segmentation technique prior to the feature extraction step to derive more accurate features. The other explains achieving perfect segmentation is impossible, then selects ROI manually and extracts features from the ROI area. After that, I will present my solution for feature extraction to deal with the challenge from a different perspective.

2.2.1 Background

Many feature extraction techniques and algorithms are presented for breast lesion detection, segmentation, staging, or prediction in digital mammograms [74, 75]. Earlier studies usually compute global image features from the entire breast areas segmented from the mammograms to predict/detect cancer [76-79]. In other words, for the image processing step since an accurate segmentation, because of the chest wall, different sizes of breasts, and the other artifacts, is hard to get, the studies preferred to extract global image features for their analyses. The study in [79] is based on dynamic contrast-enhanced breast magnetic resonance imaging (DCE-MRI). While this study presents a global feature extraction technique, they prefer DCE-MRI for the imaging modality since more subtle cancers are possible to detect in this imaging modality in comparison with mammography. The contract's enhanced characteristic of this imaging modality helps them bypass the segmentation step and overcome this challenge. However, in order to investigate whether applying the local features computed

from mammograms can improve the accuracy of predicting breast cancer risk, the researchers in [18] present a new idea for local image feature extraction that is briefly described below.

In this approach, the entire breast area is first segmented into nine strip-based local regions. CAD scheme computes both global- and local-based bilateral asymmetrical image features. An initial feature pool includes 190 features. It contains the features related to the spatial distribution and structural similarity of grayscale values computed from the original images and the features related to the magnitude and phase responses calculated from the images being filtered by multidirectional Gabor filters and difference of Gaussian filters. An overall view of this method is shown in Figure 2-1.

For the feature extraction domain, the authors first apply a CAD scheme to distinguish the nipple as referenced point and the breast skin-line, employing the available methods in [80, 81]. The CAD scheme then segments the global breast region and several local regions, including horizontal strips, vertical strips, and DoG (difference of Gaussian) primary element regions. Specifically, the global breast region is derived by extracting the largest rectangles enclosed in the whole breast region. For strips, nine horizontal strips are segmented. Nine vertical strips are also segmented parallel to the chest wall. Then, the DoG basic element regions are segmented with DoG filters and a morphology approach.

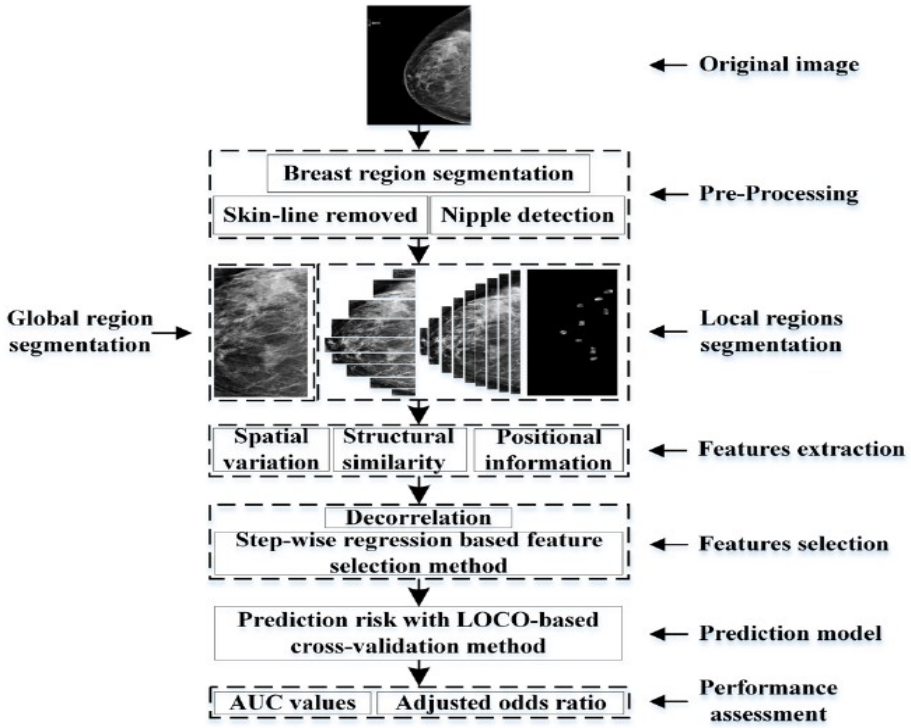


Figure 2-1Flowchart of the procedure for short-term breast cancer risk [18].

This study shows that local image features relayed on a perfect segmentation technique yield higher performance than global features, and image features extracted from the global and the matched local regions are complementary (not highly correlated). Although the method of extracting local features, and combining them with global features is impressive, and the results show it can significantly increase performance, the AUCs are still not high enough for clinical practice. This paper also uses the method in [81] for nipple detection, which is the base of this paper's segmentation step. This method for nipple detection is not accurate enough in challenging conditions. Hence, achieving a precise segmentation may not be possible for these conditions. Accordingly, in recent years, multi-classes feature extraction techniques from local areas have been explored to aid in ultrasound imaging analysis and help the radiologist concentrate multiple frames for examinations instead of one pre-selected frame. In [82], the authors discussed developing a scheme based on local feature extraction techniques to assist radiologists in diagnosing breast cancer based on ultrasound imaging. In this study, they prefer

local feature extraction because of global artifacts, while automatic ROI segmentation, especially for ultrasound images, because of low contrast, is hard to get, and handcraft segmentation should be a solution to present generally accepted methods. Despite great research efforts on the challenge of feature extraction, ML-based CADx schemes of mammograms have not been accepted and used in clinical practice. It still faces multiple technical challenges to improve CADx performance and robustness. For instance, as discussed, due to the imaging system's quality (like ultrasound images) or the overlap of dense fibro-glandular tissues on mammograms, accurate lesion segmentation is often tricky and unreliable, which can substantially affect the performance and robustness of CADx schemes [10].

2.2.2 Hypothesis and Proposed Approach

To deal with global and local feature extraction objections and extract practical global features and find local lesions patterns to drive local features from them, I developed a CADx scheme that focuses on the frequency-domain characteristic for global analyses. In contrast, it forms a tree-based structural SSIM feature extraction with no need for an initial segmentation procedure, when it is about lesion investigations. In my method, I will explain that under frequency domain analyses for global feature extraction, it is possible to overcome the adverse effects of artifacts on the final feature vector as much as possible. Besides, with the tree-based structural method for SSIM calculation, the chance of segmentation error rate on the local analyses will be reduced. The premise of my idea is that by avoiding difficulty and errors in lesion segmentation, it is possible to develop and apply the global image feature analysis schemes in the CAD-related quantitative image informatics field, which has the potential to be more efficient and robust [83].

In this study, I hypothesized that global image feature analysis schemes could be developed and applied to predict the likelihood of cases being malignant if the radiologists detect the suspicious lesions (i.e., soft tissue-based masses) on mammograms. Thus, I proposed

investigating a new CADx scheme with two unique approaches—first, global image feature extraction. Second, taking advantage of bilateral asymmetrical characteristics computed from left and right breasts. If successful, this new approach may be able to provide radiologists a new CADx-generated image marker or risk prediction score to support their decision-making in classifying malignant and benign lesions to increase diagnostic accuracy. Thus, this study's objective is to test my hypothesis using a relatively large and diverse digital mammography image dataset. Compared to already available studies, this study will demonstrate the feasibility of developing a new global image feature analysis based CADx scheme of mammograms with high performance. This new CADx approach will be more efficient in development and potentially more robust in future applications by avoiding difficulty and possible breast lesion segmentation errors.

This study's unique contribution is that we demonstrate the feasibility of developing a new image feature analysis-based CADx scheme to predict the likelihood of mammography cases being malignant without lesion segmentation. The details of this study will be in chapter three.

2.3 Feature optimization/selection and their impacts on CAD performance

In chapter two, I presented the standard feature optimization methods and the logic behind their functional procedure. Then, in part 2.1, I explained that many parameters are involved in the process of feature optimization, and achieving an optimal feature-set out of initial features is still a challenging trial. Researchers commonly use feature evaluation techniques to develop the most prominent group of the initial extracted features to achieve the best performance of ML embedded CAD schemes of medical images. In this part, I will present some of the already available research and studies on this step, then propose two different hypotheses as prominent solutions to handle it well.

2.3.1 Background

T-test and F-test are two statistical feature selection methods utilized in the medical imaging area like the other areas to derive relevant features from the initial feature matrix. Authors in [84] propose a mammogram classification scheme to classify the breast tissues as normal, benign, or malignant based on these feature selection methods. To this aim, first, a feature matrix is formed employing GLCM to all the complex components from 2D DWT of the ROI of mammography images.

Then, they utilize t-test and f-test on the feature space separately. The relevant features are used to train a backpropagation neural network (BPNN) classifier for classification. It is shown in this paper that using the image features selected by the t-test approach outperforms using the features determined by the f-test method. They also compare their method with another scheme in [85] based on PCA, which the results show t-test exceeds PCA as well. Another study in [86] proposed a scheme for the diagnosis of breast cancer from mammogram images. In this study, the T-test and F-test performances are compared, and the results show based on the provided matrix of features, size of the feature-set, and initial combination of the features, they can present superior outcomes to each other.

The negative point about the t-test is that t-test fails when there is a small number of cases in the dataset, or the estimated variances are not similar between different classes. However, this study's experiments and results show the importance of the feature selection method and proper machine learning scheme for a classification problem.

Another scheme based on the filter-based feature selection method is presented in [39] to promote early detection of Alzheimer's. The main idea is based on filtering the best combination of initial features on voxels and then filtering the best combination of healthy or Alzheimer subjects' voxels in a supervised manner. Like the other methods, segmentation is done for Volumes-Of-Interest (VOIs), and statistical moments are extracted as features. Then, a

supervised filter-based feature selection process is developed. It enables to pick the best feature subset and rank them according to the (AUC) values.

To validate the effectiveness of this feature selection method, the authors perform experiments using a comparison with the state-of-art of Feature Selection methods (FS), including Student's t-Test analysis [87], Fisher score [88], Support Vector Machine feature elimination (SVM-RFE) [89], feature selection with Random Forest [88], minimum Redundancy Maximum Relevance (mRMR) [90] and ReliefF [91]. The Voxel-Based Analysis (VBA), considered as a baseline classification approach, is used for comparison purposes. Experiments show the method in this paper could outperform all other feature selection techniques.

While the scheme in [39] outperforms different feature selection techniques, filter-based methods fail to model the relationship between the features, which can effectively boost the performance. On the other hand, embedded methods can overcome this weakness by modeling possible features' relationships and acting as a solution to this obstacle. As an embedded technique for feature selection, [14] proposes a new feature selection method based on Fisher criterion and genetic optimization, called FIG for short, to tackle the common CT imaging signs of lung diseases (CISL) recognition problem. In the FIG feature selection method, the Fisher criterion is applied to evaluate feature subsets, based on which a genetic optimization algorithm is developed to find out an optimal feature subset from the candidate features. The researchers use the FIG method to select the CISL recognition features from various types of features, including bag-of-visual-words [92] based on the histogram of oriented gradients, the wavelet transform-based features, the local binary pattern, and the CT value histogram. Then, the selected features cooperate with each of five commonly used classifiers, including support vector machine (SVM), Bagging (Bag), Naive Bayes (NB), k-nearest neighbor (k-NN), and AdaBoost (Ada) to classify the regions of interests (ROIs) in lung CT images into the CISL

categories. For all the considered classifiers, the FIG method brought better recognition performance than the full set of original features and any single type of feature.

2.3.2 Hypothesis and Proposed Approach

Based on the idea of embedded methods, I developed two different CAD schemes established for feature optimization, which are not just based on removing less functional features, but also on considering the most practical combination of features for the purpose that we have. In other words, in these two studies comparing to all previous studies, I try to regenerate the best group of final features out of the initial feature vector. In other words, the main contributions of my studies are not about features selection; they are about optimization of the features as far as possible either by regenerating a new group of features out of the initial feature vector, or by considering the fact of more robust and reliable classification under preserving the distance of the points in lower dimension spaces of features. Two different tasks are supposed to evaluate the performance of my proposed schemes. The first one is on the application of predicting the risk of cancer in the next sequence of mammography images. The second technique is on the assignment of classifying benign cases from malignant ones.

For the first task, in order to automatically identify a set of compelling mammographic image features and build an optimal breast cancer risk stratification model, I investigate the advantages of applying a machine learning approach embedded with a locally preserving projection (LPP) based feature combination and regeneration algorithm as a prerequisite to predict short-term breast cancer risk.

This study aims to develop and apply a new machine learning approach to create a small effective feature vector to build an optimal machine learning classifier using a relatively small training image dataset. Inspired by the deep learning technology, which directly uses input images to generate an efficient feature vector for classification [93], I propose to apply a locally preserving projection (LPP) based feature combination algorithm [94] to reduce the

dimensionality of feature space and then build a new image features based short-term breast cancer risk prediction model. Unlike the conventional feature selection methods that filter and select a set of existing optimal features from the initial feature pools, LPP generates a new optimal feature vector involving components that are different from any original features in the existing feature pool. But like embedded methods, it can overcome the weakness of feature dependence and correlation by modeling possible feature relationships. In this study, I demonstrate that applying the LPP algorithm effectively reduces feature dimensionality and yields higher and potentially more robust performance in predicting short-term breast cancer risk [95].

In this study, I mainly focus on dimensional reduction with regeneration techniques on the initial feature set inspired by deep learning techniques. Results of this analysis show significance of this stage on any ML-based CAD scheme. The details of this study will be presented in chapter four.

For the second study on feature optimization, to better address this challenge and more reliably regenerate image feature vector for developing CAD schemes of medical images, I investigate and test another feature regeneration method namely, a random projection algorithm (RPA), which is an efficient way to map features into a space with a lower-dimensional subspace, while preserving the distances between points under better contrast [96]. In fact, since RPA has been tested and implemented in a wide range of engineering applications, including handwriting recognition [97], face recognition and detection [98], visual object tracking and recognition [99, 100], and car detection [101]. Thus, motivated by the success of applying RPA to the complex and nonlinear feature data used in many engineering application domains, I hypothesize that RPA also has advantages when applying to medical images with heterogeneous feature distributions. To test my hypothesis, I conduct this study to investigate feasibility and potential advantages of applying RPA to build optimal feature vector and train ML model implemented in a new CAD scheme to classify between malignant and benign breast

lesions depicting on digital mammograms. This study demonstrates that the random project algorithm is a promising method to generate optimal feature vectors to help improve the performance of machine learning models of medical images. Details of this method is available in chapter five.

2.4 Application of Deep Learning for Optimal Feature Extraction and Classification

In recent developments, deep learning technology based on CNN's becomes the state of the artificially intelligent medical imaging approach for all the areas covered by classical machine learning schemes. We can take advantage of deep learning strategies to bypass image processing and handcraft feature extraction steps and put all the related challenges behind. Strictly speaking, DL algorithms extract automatic features from the original images under different iterations in an end-to-end way. It is possible to apply them directly to a Softmax classifier, or a classical machine learning scheme like SVM, or optimize them first with one of the previously mentioned feature optimization techniques and then apply them to a classification scheme [102]. In this section, I will present some already available DL-based research and studies on the area of medical imaging and then present my hypothesis as a practical way to utilize DL-based algorithms and handle some data-related challenges in the area of medical imaging.

2.4.1 Background

There are many available types of CNN-based CAD schemes in the area of medical imaging for many purposes like diagnosis, detection, classification, or segmentation of disease patterns [102-108]. For instance, in [106] which is presented for malignant and benign classification, ROIs of mammogram images are fed into a modified version of AlexNet for classification purposes. The results are compared with GoogleNet as a more advanced version of CNNs. The simulations show the superiority of AlexNet in this classification. The results show although GoogleNet is a more updated version of CNN's compared to AlexNet, that does

not mean it will always make better results. In another study in [108], a fully integrated CAD diagnosis scheme is presented to detect, segment, and classify lesions. In other words, a You-Look-Only-Once (YOLO) [109] approach is utilized for mass detection, a full resolution convolutional network (FrCN) is used for mass segmentation. Finally, a CNN is employed for the classification step. While it is a fully automatic diagnosis process, three stages with different CNNs make the overall system complicated and the training process so time-consuming.

There is a recent publication that is a fascinating study based on a combination of feature extraction, feature optimization, and ensemble learning. Ensemble learning is a process in which better predictive performance is obtained by combining the results from multiple classification models into one high-quality classifier. This is another way of optimizing outcomes by error compensation out of different classifiers. In [105], a new method for classifying medical images is introduced that uses an ensemble of different CNN architectures. Since different CNN architectures learn different semantic image representation levels, an ensemble of CNNs will enable higher quality features appropriate for classification. In this paper, the feature extraction scheme is CNN-based. A PCA optimization technique is applied to them to make them more efficient, and classical classifiers are used to extract the final results. An overview of this ensemble learning is shown in Figure 2-2.

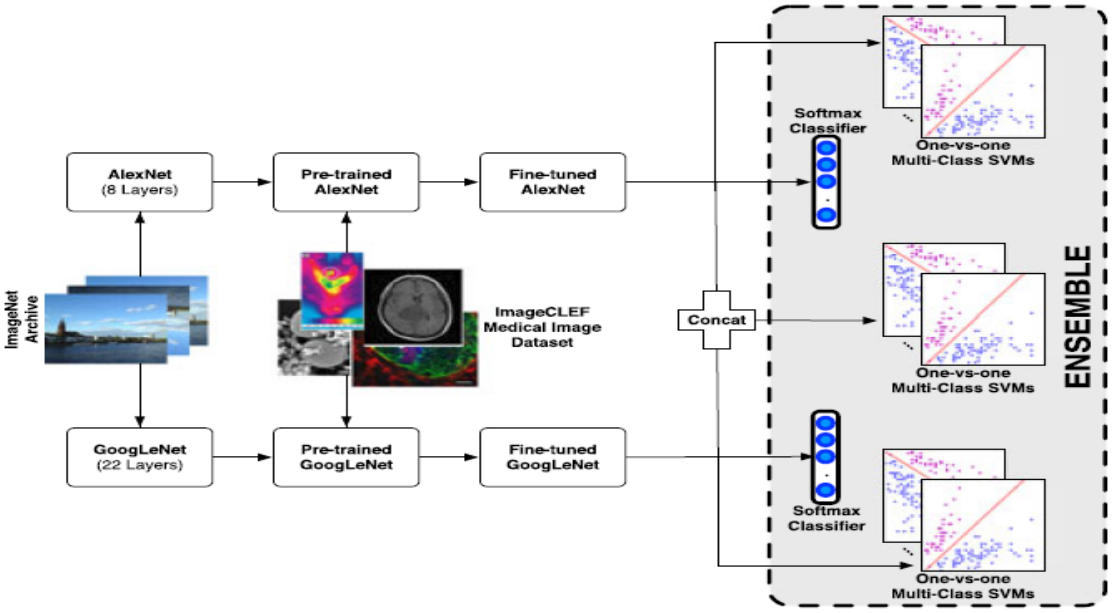


Figure 2-2 Overview of an ensemble method [67].

As this figure shows, two CNN architectures (AlexNet [41], and GoogleNet [110]) that has been pre-trained (initialized) on natural image data are fine-tuned first. Then, each of the fine-tuned CNNs is utilized in two ways: 1) as an image feature extractor with the independent feature vectors concatenated to train multiclass SVMs, and 2) as a classifier forming SoftMax probabilities. The following probabilities from the SVMs and SoftMax classifiers' ensemble are prepared to determine the image's class.

From AlexNet, 4096, and GoogleNet, 1024 features are extracted using the activations of the last fully connected layer of the fine-tuned network. For efficient classifier training, the dimensionality of AlexNet features is reduced to 459, and the dimensionality of GoogleNet features is reduced to 108 using PCA. SVM and Softmax classifiers are applied to each of these PCA's output features separately, and SVM is applied to a concatenation of these features that have 508-D. Then the posterior probability of the whole scheme is defined based on all these five classifiers' outputs. This study's experimental results present this method outperforms techniques that simply utilize ResNet, GoogleNet, and VGGNet for classification purposes.

Another similar ensembling technique is developed in [107] for the classification of mammograms. It used five different states of the art CNNs for the feature extraction phase and a neural network for the classification phase. The difference with the previous study is that in this study, instead of using PCA for feature optimization, the concatenated features are applied to a neural network for overall optimization. Hence, the neural network

Results of both these researches show even on deep learning schemes that are entirely automatic for classification in the medical imaging informatics field and have a high level of performance on classification; we can either come up with new ideas on reducing features-dimensionality with data reduction and feature selection techniques like PCA, and use classical machine learning schemes like SVM to combine results of these schemes to achieve better results than baseline methods in the area, or utilize a neural network for an optimal segmentation, learning and classification purpose.

2.4.2 Hypothesis and Proposed Approach

Despite the promising results reported in previous studies, many issues have not been well investigated and addressed regarding how to train deep learning models optimally. For instance, practical solutions for the challenge of small datasets, and imbalanced data in medical imaging since usually the number of patients with the disease is much less than normal cases. Besides, whether applying image preprocessing algorithms can improve the performance and robustness of the deep learning models. To better address some of the challenges or technical issues, I develop and test a new DL-based CAD scheme of chest X-ray radiography images. The hypothesis in this study is that instead of directly using the original chest X-ray images to train deep learning models, it is practical to apply image processing algorithms to remove the majority of diaphragm regions, normalize image contrast and reduce image noise, and generate a pseudocolor image based on them to feed in 3 input channels of the existing deep learning models that are pre-trained using color (RGB) images in the transfer learning process.

The premise to present the CNN-based technique for Covid-19 classification is based on the idea that, like posterior processing as a way of improving the performance of baseline CNNs, it is possible if we work on preprocessing steps and enhance the quality of the images before applying them to the CNNs. The study will demonstrate that adding two image preprocessing steps and generating a pseudo-color image plays an essential role in developing deep learning CAD schemes. Furthermore, I compare the proposed method with other already available methods as baselines to show my approach's superiority and the effectiveness of pre-processing techniques on CNN-based algorithms [111]. The details of the study will be presented in chapter six.

3 Development and Assessment of a New Global Mammographic Image Feature Analysis Scheme to Predict Likelihood of Malignant Cases

3.1 Introduction

Mammography is the only clinically acceptable imaging modality to detect breast cancer in current population-based breast cancer screening [112, 113]. Due to the quite low cancer detection yield (i.e., detecting 3.6 cancers per 1000 (0.36%) mammography screenings [114, 115]) and higher recall rate (i.e., ~10%) in breast cancer screening environment, reading and interpreting screening mammograms is difficult and time-consuming for radiologists [116]. To assist radiologists more accurately and efficiently reading mammograms and reducing inter-reader variability, computer-aided detection (CADe) schemes of mammograms have been developed and used in the clinical practice as "a second reader" for the last two decades [1]. Although previous observer performance studies reported that using CADe might help radiologists detect more cancers that may be previously missed or overlooked by radiologists (i.e., [2]), the clinical data analysis studies showed that using CADe increased false-positive recalls and reduced radiologists' performance measured by areas under the receiver operating characteristics curves (i.e., [3]). Thus, the specificity of current mammographic imaging remains lower in clinical practice. Approximately only one in four lesion biopsies are proved to be malignant [117]. The higher false-positive recall rates add anxiety with potentially long-term psychosocial consequences [4] and physical harm to many cancer-free women who participate in mammography screening due to cumulative x-ray radiation and unnecessary biopsies [5]. The high false-positive recall rates also associate with a high economic burden on the healthcare system [6, 118]; thus, in order to help improve the efficacy of mammography screening, developing the computer-aided diagnosis (CADx) schemes that aim to assist

radiologists in their decision-making to better assess the risk of the detected suspicious breast lesions being malignant and reduce the unnecessary biopsies of benign lesions have been attracted broad research interest for the last two decades [7, 119].

Despite great research effort, CADx schemes of mammograms have not been accepted and used in clinical practice. It still faces multiple technical challenges to improve CADx performance and robustness. For example, previous schemes typically include 3 steps which (1) apply image processing algorithms to segment suspicious lesions depicting on mammograms, (2) compute images features from the segmented regions, and (3) train multi-feature fusion-based machine learning classifiers [8]. However, due to the overlap of dense fibro-glandular tissue on mammograms, accurate lesion segmentation is often difficult and unreliable, which can substantially affect the performance and robustness of CADx schemes [10]. To overcome this difficulty, researchers recently investigated and applied deep learning techniques to develop CADx schemes without lesion segmentation and hand-crafted feature computation [12, 13]. Although DL approach can avoid difficulty in lesion segmentation and manually defining image features, it requires a large and diverse image dataset to train the scheme to minimize the risk of overfitting and validate its performance, which is another difficult task.

To address these challenges, we recently investigated the feasibility of developing new computer-aided quantitative image feature analysis schemes or machine learning models based on the global mammographic image features to predict the risk of developing breast cancer in a short-term [76, 95] or risk of depicting suspicious lesions on mammograms [120]. Similar global image feature analysis schemes can also be developed using different imaging modalities to predict other clinical outcomes, such as the response of breast cancer patients to neoadjuvant chemotherapies using breast MRI [121] and the response of ovarian cancer patients to chemotherapy using CT images [122]. By avoiding difficulty and errors in lesion segmentation, our studies have demonstrated the advantages of developing and applying the global image

feature analysis schemes in the CAD-related quantitative image informatics field, which has the potential to be more efficient and robust.

In this study, we hypothesized that similar global image feature analysis schemes could be developed and applied to predict the likelihood of cases being malignant if the suspicious lesions (i.e., soft tissue-based masses) are detected by the radiologists on mammograms. Thus, we proposed to investigate a new CADx scheme with 2 unique approaches. First, the new CADx scheme identifies and selects image features computed from the entire breast area depicting on the mammograms of left and right breasts. This global approach is different from previous local region or lesion-based CADx schemes that either require lesion segmentation or define regions of interest (ROI) with the fixed size to cover the suspicious lesions. Second, the new CADx scheme uses the bilateral asymmetrical image features computed from left and right breasts. As a result, unlike the conventional single-image-based CADx scheme, this is a multi-image fusion-based CADx scheme. Although this is a new approach, the advantages of developing multi-image fusion-based CADe schemes over the single-image-based CADe schemes have been demonstrated in previous studies (i.e.,[123]). If successful, this new approach may enable to provide radiologists a new CADx-generated image marker or risk prediction score to support their decision-making in classifying between malignant and benign lesions to increase diagnostic accuracy (including reduction of false-positive recalls and unnecessary biopsies of benign lesions). Thus, the objective of this study is to test our hypothesis using a relatively large and diverse digital mammography image dataset.

3.2 Materials and Methods

3.2.1 Image Dataset

We retrospectively assembled a full-field digital mammography (FFDM) image dataset, which involves the fully anonymized images acquired from 1,959 patients who underwent routine annual mammography screening with age ranging from 35 to 80 years old. In these

patients, suspicious lesions were detected by radiologists in the original mammogram reading and diagnosis. All detected suspicious lesions were recommended, and performed biopsy. Based on the histopathology examinations of the biopsy-extracted lesion specimens, 737 cases were confirmed as positive for cancer, while other 1,222 cases had biopsy-approved benign masses.

Table 3-1 Case numbers and percentage distribution of patients' age and mammographic density rate by radiologists using BIRADS guidelines.

	Subgroup	Malignant Cases	Benign Cases
Density BIRADS			
	1	39 (5.3%)	58 (4.7%)
	2	286 (38.8%)	412 (33.7%)
	3	401 (54.4%)	702 (57.4%)
p-value = 0.878	4	11 (1.5%)	50 (4.1%)
Age of Patients (years old)			
	$A < 40$	25 (3.4%)	50 (4.1%)
	$40 \leq A < 50$	141(19.2%)	561(45.9%)
	$50 \leq A < 60$	189(25.6%)	335(27.4%)
	$60 \leq A < 70$	180(24.4%)	187(15.3%)
	$70 \leq A$	202(27.4%)	89(7.3%)

Each mammography case involves 4 images of craniocaudal (CC) and mediolateral oblique (MLO) view of the left and right breasts. The original FFDM images have a pixel size of $70\mu m$. Like the conventional CAD schemes of mammograms, all images were subsampled using a pixel averaging method with a 5×5 pixel frame to make the image size of 818×666 pixels and 12-bit pixel depth. Each pixel size is increased to $0.35mm$ [124]. Table 3-1 summarizes and compares case distribution of patient age and mammographic density rated by radiologists using BIRADS guidelines. The patients in benign groups are relatively younger than those in malignant groups, but there is no significant difference in BIRADS density rating (p=0.878).

3.2.2 Background of Image Features

After segmenting the breast area from the surrounding area region depicting on each mammogram [95], we applied a computerized scheme to extract and compute global image

features from the original mammograms in the spatial domain and the transformed maps in the frequency domain. Specifically, the feature extraction algorithm relies on the basic fact that mammography images are highly structured, which means their pixels exhibit strong dependences. Under the presence of cancer, the pixel dependency would change not only in the region of lesions, but also the surrounding parenchymal tissues in the breast area. In addition, since radiologists are quite sensitive to bilateral image feature differences related to the structural information between left and right breasts in detecting suspicious lesions and distinguish malignancy cases from benign ones, we will extract and compute the global bilateral image feature difference between the left and right CC or MLO view images to build the machine learning model for predicting the risk of the cases being malignant.

From the original FFDM images, we computed image features and applied the structural similarity index (SSIM) to measure the similarity between 2 bilateral images of the left and right breasts. SSIM was originally proposed to assess image quality based on structural similarity [125]. It has been widely utilized in the medical imaging field with higher correlation to the human visual system adapted to extract structural information of images, including our previous studies (i.e., [76]). For the SSIM assessment, if we assume two nonnegative image signals like $\boldsymbol{x} = \{x_i | i = 1,2, \dots, M\}$ and $\boldsymbol{y} = \{y_i | i = 1,2, \dots, M\}$ as two patches of each image that have been aligned to each other, we can calculate SSIM index using the following equation [126]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad 3\text{-}1$$

where $\sigma_{xy} = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu_x)(y_i - \mu_y)$, $\mu_x = \frac{1}{M}\sum_{i=1}^{M} x_i$, $\sigma_x = \frac{1}{M}\sum_{i=1}^{M}(x_i - \mu_x)^2$, $\sigma_y = \frac{1}{M}\sum_{i=1}^{M}(y_i - \mu_y)^2$, and C_1, C_2 are two positive constant.

Thus, SSIM index values range between zero and one. The maximum value is achieved when the input images are identical. The more the two input images are bilaterally different to each other, the smaller the corresponding calculated SSIM index values are.

In addition, to take advantage of computer vision that can be relatively easy to acquire and analyze image features in the frequency domain, we performed two transformations to compute bilateral image feature differences in the frequency domain. Specifically, we applied discrete cosine transform (DCT), and fast Fourier transform (FFT) as two similar and complementary ways to facilitate detecting and analyzing useful image contents change across the whole image just in a small number of components. In general, under these transformations lower spatial frequency coefficients contain more information than higher frequency components.

For DCT transformation, by assuming $f(x, y)$ as an image with size $M\ by\ N$, we applied the following general equation to the image to calculate 2D DCT of the input image:

$$F(m,n) = {}^{2}/_{\sqrt[2]{MN}}\ C(m)C(n) * \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cos\left({}^{(2x+1)m\pi}/_{2M}\right) * \cos\left({}^{(2y+1)n\pi}/_{2N}\right) \quad 3\text{-}2$$

In this equation $C(m) = C(n) = {}^{1}/_{\sqrt{2}}$ for $m, n = 1$ and $C(m) = C(n) = 1$ otherwise.

DCT transforms the information contained in pixels of special domain to frequency domain. The element in the top left corner of 2D DCT matrix is the DC term and have a value that is almost always of a large magnitude, which is summation of all pixel values. On a zigzag scanning from the top left to the bottom right corner, the farther away from the DC term, it will have components with the higher frequency with the smaller corresponding magnitude [127].

FFT transformation computes the discrete Fourier transform of its input sequence. If the input image has a specific pattern, this transform can detect it in the magnitude spectrum

components. By assuming $f(x, y)$ as an image with size $M\ by\ N$, we used the following general equation to calculate 2D discrete Fourier transform of the input images:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\exp\left[2\pi i\left(\frac{xu}{M}+\frac{yv}{N}\right)\right];\ \begin{cases} u = 0, 1, \dots, M-1 \\ v = 0, 1, \dots, N-1 \end{cases} \qquad 3\text{-}3$$

where u, v are special frequencies and $|F(u,v)| = \sqrt{F_{Re}^2(u,v) + F_{Im}^2(u,v)}$ represents magnitude spectrum, which is useful to extract specific patterns [128].

Texture of each image shows important properties of distribution pattern of the fatty and fibro glandular tissues of breast. Texture also determines the local spectral or frequency content of an image. In the frequency domain, it is mostly projected to the low-frequency coefficients. On the other hand, the noise unrelated to any specific pattern (like additive noise pattern) because of their randomness nature are mostly projected to the high-frequency components. Furthermore, in [129] it has been shown that if scanning DCT coefficients in a zigzag order, the absolute DCT coefficient values are somehow correlated to each other, which means that the absolute DCT coefficient values are horizontally, vertically, and diagonally correlated to each other. Magnitude spectrums of FFT components also have the same characteristics in each local area. We will take advantage of these attributes for feature extraction.

Moreover, because of particular characteristics of benign and malignant tumors like intensity, shape, and texture [130], the structural patterns of breasts depicting benign or malignant tumors would be different. The larger or higher grade of the malignant tumor for a case is, the larger disturbance or bilateral structural differences caused by cancer manifests become more obviously. The disturbance can be enlarged in the difference of absolute value of extracted image features from the frequency coefficients of left and right breast by considering the fact that the presence of a tumor can disturb the correlation of frequency coefficients.

3.2.3 Data Preprocessing

CADx scheme applies an image preprocessing phase to the whole FFDM images. This step includes two algorithms namely, a cropping operation and an image enhancement. Cropping is just applied to the MLO view to detect and remove chest wall area, and enhancement is done on both CC and MLO views of images to remove or reduce image noise on the black background area as well as written labels.

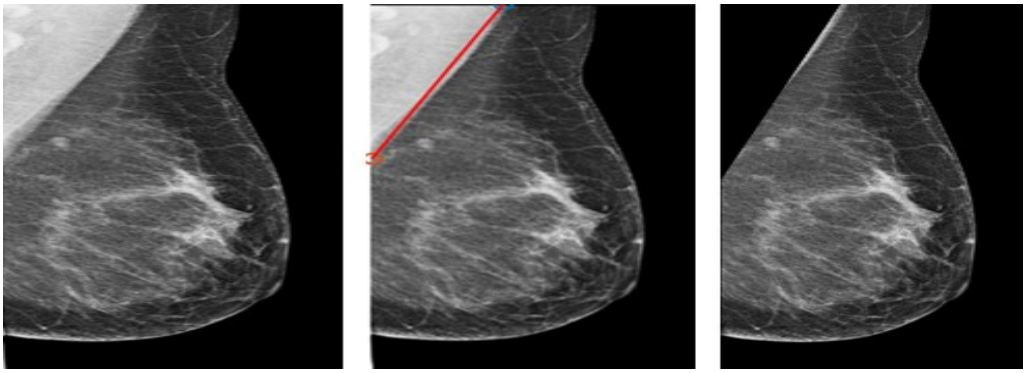


Figure 3-1 Preprocessing phase. a) the original image, b) chest wall removal step, c) denoising black area, and written labels removal.

MLO view images typically have advantages over CC view because almost all the breast area is visible, which means we can extract more information from this view, especially for the global feature extraction methods. The primary disadvantage of this view is that images also include chest wall and part of the pectoral muscle regions. The pectoral muscle area is typically brighter compared to the breast tissue and will have a negative effect on the extracted features. Due to the great variation or heterogeneity of mammograms in different cases, although developing automated schemes to segment pectoral muscle has been tried before (i.e., [131]), it remains difficult to achieve robust results when applying to a large and diverse dataset. In this study, we used a hand craft method to remove the pectoral muscle area. For each MLO image, two points are determined at the margin of the chest wall, then by plotting a straight line between these two points, the chest wall is determined, and the pectoral muscle region is deleted to select the remaining breast region for further analysis. Other non-breast areas (i.e., labels)

are also automatically deleted. An example of this preprocessing phase is given in Figure 3-1. After the pre-processing phase, all images in the dataset are saved in Portable Network Graphics (PNG) format of a lossless mode for the feature extraction phase.

3.2.4 Image Feature Extraction

After image pre-processing, the computer-aided scheme is applied to extract and compute relevant image features from the entire breast area segmented on FFDM images. These features are divided into 3 subgroups from both spatial domain and frequency domain. First, from the original FFDM images (spatial domain), the scheme computes SSIM-related features of left and right images of CC or MLO view in a tree structural shape base, which is inspired by the commonly used hierarchical methods in video data processing area for motion estimation purpose [132]. Since each FFDM image has an original size of 818 × 666 pixels, each left and right image is first divided into 4 sub-blocks with a size of 409 × 333 pixels each. SSIM is computed using Equation (3-1) for all pairs of 4 sub-blocks in the matched position of the left and right breast. The sub-block with the smallest SSIM value, which means the highest bilateral asymmetry among these 4 pairs of the matched sub-blocks, is selected. Next, the scheme continues to divide the selected sub-blocks into 4 sub-blocks again with a size of 205 × 167 pixels each. Four SSIM indices are computed and the new sub-block with the smallest SSIM index value is selected again. Such a process is repeated 6 times or iterations. In the last iteration, the size of the sub-block is reduced to 13 × 11 pixels. From these 6 iterations, the scheme selects 6 SSIM index values representing the highest bilateral asymmetry of breast tissue patterns with gradually deceased sub-block size. Figure 3-2 illustrates a block diagram of this process.

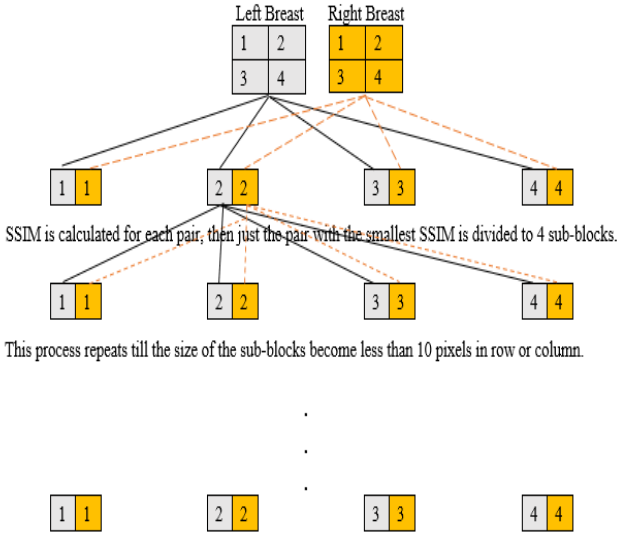


Figure 3-2 Block diagram of the proposed method for SSIM feature extraction.

In computing SSIM index value, several parameters need to be determined by the experiments. Based on our experimental results in computing SSIM of bilateral FFDM images for cancer risk assessment [76], the default parameter values are set up as 0.05 for constants C_1 and C_2, and 8 for window size used in Equation (3-1). Additionally, due to the heterogeneity of clinical cases (i.e., the variation lesion size and surround parenchymal tissues), it is not possible to predetermine an optimal sub-block size to compute SSIM index. Thus, in this study, we selected all 6 smallest SSIM index values computed in the above iterations to build an SSIM feature pool or vector.

Second, after DCT and FFT transformation, the scheme computes 2 two-dimensional (2D) DCT and FFT matrixes of the whole input image using Equations (2) and (3), respectively. Hence, each image has a 2D matrix of DCT coefficients and a 2D matrix of FFT coefficients. By filtering out the last 10 percent of high-frequency components, the redundant information is mostly filtered out with respect to the information related to the pattern of the breast. In this way, frequency domain coefficients are more suitable for feature extraction rather than pixel domain coefficients [133].

After preprocessing on the frequency coefficients, the 2D matrixes are changed to row format to reduce the computational complexity of the feature extraction phase. Hence, a sequence like $X = (x_1, x_2, \dots, x_K)$ represents these coefficients in row format. Then, the following features are extracted and computed.

From the DCT and FFT frequency domain, our scheme computed the following statistical moments-related features. Based on [134], by assuming that a sequence like $X = (x_1, x_2, \dots, x_N)$ is a finite population of size N, the scheme can compute an unknown probability density function (PDF) *p(x)* for this targeted population. The nth row moment for this population is given by:

$$\widehat{m}_n = \frac{1}{N}\sum_{i=1}^{N} x_i^n \qquad \textit{3-4}$$

where the 1st row moment ($n = 1$) is the mean (μ) of this population. By centralizing this equation, the scheme calculates the next centralized momentums for the population with:

$$\widehat{m}_n = \sum_{i=1}^{N} p_i (x_i - \mu)^n \qquad \textit{3-5}$$

That is an unbiased estimate of nth moment:

$$m_n = EX^n = \int_{-\infty}^{\infty} p(x) x^n dx \qquad \textit{3-6}$$

According to Equation (3-6), $p(x)$ is weighted by x^n, so that any change in the *p(x)* is polynomially reinforced in the statistical moments. Thus, by considering DCT and FFT components as finite populations of the input images, any changes in their PDF due to the presence of malignant lesions is polynomially reinforced in the statistical moments of the computed coefficients. In this study, we utilized the statistical moments to catch bilateral image

feature differences in both DCT and FFT maps of left and right breasts. Using Equation (3-4), the scheme computes the mean of the frequency components, and using Equation (3-5) for $n = 2,3,4$, the scheme computes variance, skewness, and kurtosis of the frequency components. Additionally, the scheme also computes other popular statistical features, including entropy, correlation, energy, root mean square level, uniformity, max, min, median, range, and mean absolute deviation from the DCT and FFT maps. Then, the absolute differences of these matched image features from the left and right view maps are computed to represent global bilateral differences of the left and right breasts in DCT and FFT-based frequency domains. Table 3-2 also lists the 14 features computed from DCT and FFT maps.

Table 3-2 The computed SSIM, DCT, and FFT image Features.

Feature category	Feature Description
SSIM features computed from original FFDM images	Six SSIM indices computed using Equation (3-1) from the six pairs of sub-blocks with the gradually reduced size.
Features computed from frequency domain of DCT and FFT transformed maps	1.Mean, 2. variance, 3. skewness, 4. kurtosis, 5. entropy, 6. correlation, 7. energy, 8. root mean square level, 9. uniformity, 10. max, 11. min, 12. median, 13. range, 14. mean absolute deviation

In summary, our scheme computes 34 features from two bilateral images or maps of the left and right breasts (as shown in Table 3-2). Since each case has two sets of bilateral images acquired from CC and MLO view, the total computed image features are 68.

Figure 3-3 shows a schematic diagram of the feature extraction phase to show how the scheme extracts each sub-group of features from each of 4 individual images (LCC, RCC, LML, RML) of a case and combine them to create the final feature vector (F_{fusion}) of 34 features in each of CC or MLO views. Specifically, $lccF_{dct}$ and $lccF_{fft}$ are DCT and FFT features computed from CC view image of left breast, while $rccF_{dct}$, $rccF_{fft}$ are DCT and FFT features computed from the CC view image of the right breast. Similarly, $lmlF_{dct}$, $lmlF_{fft}$ and $rmlF_{dct}$, $rmlF_{fft}$ are DCT and FFT features computed from MLO view images of left and right breast, respectively. Last, $F_{ssim}cc$ is vector of SSIM features related to two bilateral CC view images

(LCC, and RCC), and $F_{ssim}ml$ is vector of SSIM features related to two bilateral MLO view images (LML and RML).

After computing these 34 image features from the bilateral images of one view, we computed and generated 2 correlation matrices for CC and MLO view (Figure 3-4). The results indicate that majority of these features are not highly correlated (i.e., $r < |0.25|$ as shown by the light to dark blue color in Figure 3-4), which can provide complementary information to predict the likelihood of the case being malignant.

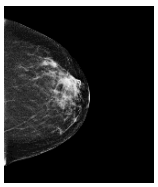
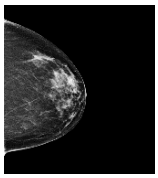
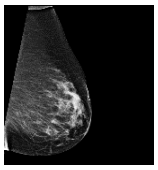
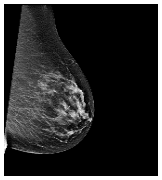
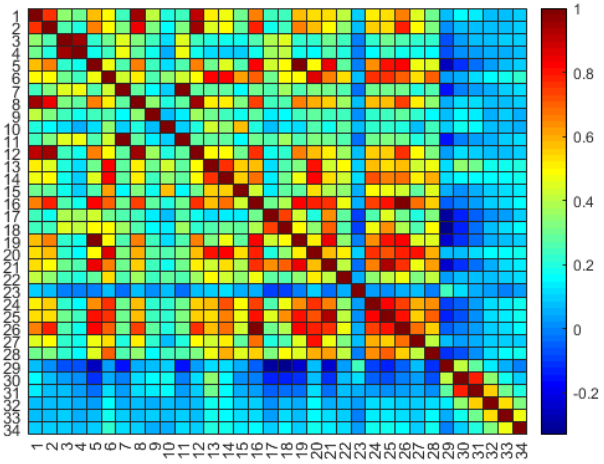
<table>
<tr>
<td rowspan="2">LCC </td>
<td>$lccF_{dct} = [f_{1c}, f_{2c}, \dots, f_{14c}]$</td>
<td rowspan="2">$F_{dct}cc = abs(lccF_{dct} - rccF_{dct})$</td>
<td rowspan="8">$F_{fusion} = [F_{dct}cc, F_{dct}ml, F_{fft}cc, F_{fft}ml, F_{ssim}cc, F_{ssim}ml]$</td>
</tr>
<tr>
<td>$lccF_{fft} = [f_{1c}, f_{2c}, \dots, f_{14c}]$</td>
</tr>
<tr>
<td rowspan="3">RCC </td>
<td rowspan="2">$F_{ssim}cc = [f_{1c}, f_{2c}, \dots, f_{6c}]$</td>
<td>$F_{fft}cc = abs(lccF_{fft} - rccF_{fft})$</td>
</tr>
<tr>
<td rowspan="2">$F_{ssim}cc = [f_{1c}, f_{2c}, \dots, f_{6c}]$</td>
</tr>
<tr>
<td>$rccF_{dct} = [f_{1c}', f_{2c}', \dots, f_{14c}']$</td>
</tr>
<tr>
<td rowspan="2">LML </td>
<td>$rccF_{fft} = [f_{1c}', f_{2c}', \dots, f_{14c}']$</td>
<td rowspan="2">$F_{dct}ml = abs(lmlF_{dct} - rmlF_{dct})$</td>
</tr>
<tr>
<td>$lmlF_{dct} = [f_{1m}, f_{2m}, \dots, f_{14m}]$</td>
</tr>
<tr>
<td rowspan="3">RML </td>
<td rowspan="2">$F_{ssim}ml = [f_{1m}, f_{2m}, \dots, f_{6m}]$</td>
<td>$F_{fft}ml = abs(lmlF_{fft} - rmlF_{fft})$</td>
</tr>
<tr>
<td rowspan="2">$F_{ssim}ml = [f_{1m}, f_{2m}, \dots, f_{6m}]$</td>
</tr>
<tr>
<td>$rmlF_{dct} = [f_{1m}', f_{2m}', \dots, f_{14m}']$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$rmlF_{fft} = [f_{1m}', f_{2m}', \dots, f_{14m}']$</td>
<td></td>
</tr>
</table>

Figure 3-3 Feature extraction phase of the proposed method.

3.2.5 Classification Phase

In this phase, we built multiple feature fusion-based machine learning models to predict the likelihood of the cases being malignant. Although many different types of machine learning classifiers (i.e., artificial neural network, Bayesian belief network, and logistic regression model) can be used for this purpose, based on our previous experience in developing the variety

of CAD schemes of medical images, we chose to train and build support vector machine (SVM) based machine learning models to predict the likelihood of the cases being malignant. To achieve high robustness, a popular RBF kernel was selected to build the SVM model, which has demonstrated good performance and low computational cost in our previous studies [135, 136]. Specifically, in each CC or MLO view images, we built 4 SVM models using image features computed from (1) the original FFDM image (6 SSIM based features), (2) DCT maps (14 features), (3) FFT maps (14 features), and fusion of 34 features computed. After comparing the performance of the SVMs trained using only one view images, we also fuse the image features computed from the two view images to retrain and test 4 new SVM models.



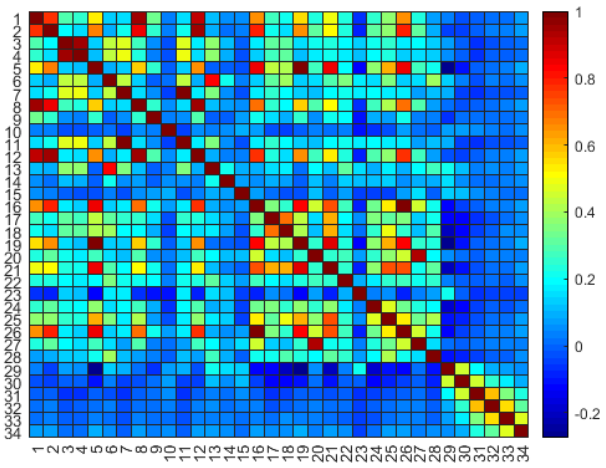


Figure 3-4 Correlation coefficient matrices of 34 image features computed from CC (top) and MLO (bottom) view images.

Each SVM-based prediction model is applied to the entire image dataset of 1,959 cases to predict the likelihood of the cases being malignant. To train each SVM and assess its performance, we applied a 10-fold cross-validation method. The SVM model produces likelihood or prediction scores ranging from 0 to 1 in the testing phase. The higher score indicates the higher risk or likelihood of the case being malignant. Using the prediction scores computed from all 1,959 cases, a receiver operating characteristic (ROC) curve is generated and the area under the ROC curve (AUC value) is computed as an evaluation index.

Then, to evaluate an absolute classification accuracy for the proposed scheme, we also applied an operating threshold (T = 0.5) on the SVM-generated prediction scores. All cases are divided into two malignant and benign classes to generate a confusion matrix. From the confusion matrix, the overall prediction or classification accuracy, sensitivity, specificity and odds ratio (OR) are calculated as well. Furthermore, we sort the SVM-generated detection scores for all cases in an ascending order and select 5 threshold values to segment all cases into 5 sub-groups. Then, based on the available multivariate statistical model included in a statistical software package (R version 2.1.1, http://www.r-project.org), we calculated the adjusted OR values and detected the possible ORs increasing trend with the increased classification scores.

In addition, to test whether we can further reduce the dimensionality of the feature space to identify better features, we applied Principal Component Analysis (PCA) as a feature analysis and regeneration method to reduce feature vector size and train SVM models. The performance levels of the SVM models trained with and without applying the PCA method were compared. All computation tasks were conducted using the MATLAB R2019a package. Figure 3-5 illustrates a complete block diagram of the proposed scheme and testing method.

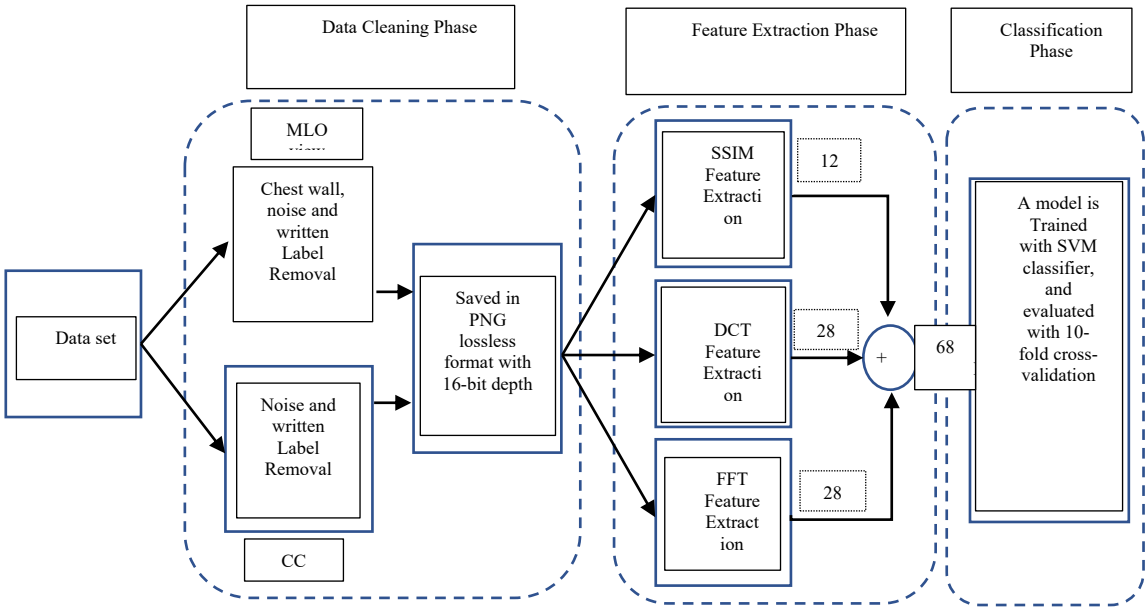


Figure 3-5 A summarized block diagram of the proposed scheme for classification of benign and malignant tissues in mammography imaging.

3.3 Results

Figure 3-6 shows CC and MLO images of one malignant and one benign case. Using the global bilateral image feature analysis, the SVM-generated prediction scores are 0.82 and 0.37 in these 2 cases, respectively. Table 3-3 shows and compares AUC values and overall classification accuracy after applying the operation threshold (T = 0.5). The results show that using the image features computed from the bilateral MLO view images yielded significantly higher performance than using image features computed from bilateral CC view images ($p < 0.05$).

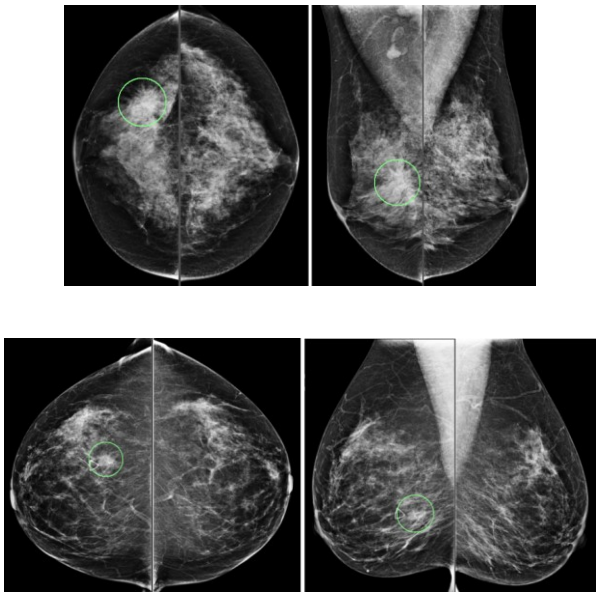


Figure 3-6 Illustration of one malignant case (the first row) and one benign case (the second row). The detected masses are circled (Green Color) in the images.

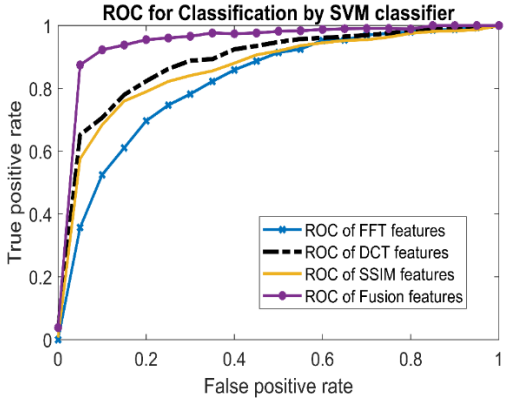


Figure 3-7 Comparison of 4 ROC curves generated by four SVMs trained using image features computed from both CC and MLO view images.

Table 3-4 summarizes and compares the computed AUC values of 4 SVM models trained using image features computed from both CC and MLO view images. Figure 3-7 shows 4 corresponding ROC curves. The results indicate that using 3 subgroups of features computed from the original FFDM images and 2 transformation maps, AUC values range from 0.85 to 0.91. After fusion of all image features computed from 3 subgroups, the AUC value of the 4th SVM model significantly increases to 0.96±0.01 (with $p<0.01$). In addition, the standard deviation after fusion of 3 subgroups of image features is also substantially decreased as

comparing to the use of one subgroup of image features, which indicates the increase of reliability of the 4th SVM model performance (AUC value).

Table 3-5 shows and compares 4 confusion matrices generated by the SVM-generated prediction scores after applying an operational threshold (T = 0.5). From these confusion matrices, additional performance indices can be computed, as shown in Table 3-6. It shows that SVM trained using the subgroup of DCT features yields the highest overall prediction accuracy as comparing to the other two SVMs trained using SSIM and FFT features. However, by fusion of all 68 features, the SVM model yields further increased overall prediction accuracy (92%).

After dividing 1,959 testing cases into 5 subgroups of an approximately equal number of cases (∼392) based on the SVM-generated prediction scores (Table 3-7), the adjusted odds ratios (OR) increased from 1.0 in the baseline subgroup with the lowest classification scores to 25,220 in the 5th subgroup with the highest prediction scores (the highest chance of being malignant). Regression analysis of the adjusted OR data also shows an increasing trend of odds ratios with the increase in SVM-generated prediction scores. The slope of the regression trend line between the adjusted ORs and SVM-generated scores is significantly different from zero slope ($p < 0.01$).

By applying a PCA algorithm to reduce feature space dimensionality, we trained and tested SVM models with the increased number of the PCA-regenerated features. The highest AUC value is 0.94, and the highest overall prediction accuracy after applying the same operation threshold of T = 0.5 is 91%, which involves 65 numeric components produced by the PCA algorithm. The performance is slightly lower than the SVM trained using all 68 features, as shown in Table 3-4 and Table 3-6.

Table 3-3 AUC and Accuracy for different sub-group of features on CC view in compare with MLO view.

Feature sub-groups	Number of features	AUC ± STD	Accuracy (%)
FFT, CC view	14	0.63 ± 0.025	66
FFT, MLO view	14	0.84 ± 0.017	77
DCT, CC view	14	0.62 ± 0.026	64
DCT, MLO view	14	0.89 ± 0.015	83
SSIM, CC view	6	0.53 ± 0.026	63
SSIM, MLO view	6	0.78 ± 0.021	71
Fusion, CC view	34	0.65 ± 0.027	67
Fusion, MLO view	34	0.94 ± 0.009	89

Table 3-4 Computed area under ROC curve using individual group of features on both CC and MLO views.

Feature sub-group	Num. of features	AUC	STD	95% CI
FFT features	28	0.85	0.018	[0.80, 0.90]
DCT features	28	0.91	0.013	[0.89, 0.95]
SSIM features	12	0.89	0.016	[0.85, 0.92]
Fusion of all features	68	0.96	0.007	[0.95, 0.97]

Moreover, to test the significance of preprocessing phase (as shown in Figure 3-1) of whether removal of pectoral muscle regions can boost the performance of the CADx scheme, we recomputed all image features from the bilateral MLO view images without removing the pectoral muscle and retrained SVM classification models. Table 3-8 compares the classification performance of the SVM models trained and tested using the image features computed from MLO view images with and without removing pectoral muscle areas depicted on the images. The result shows significant improvement of classification performance by removing the pectoral muscle areas from the MLO view images ($p<0.05$). For example, the AUC value of the SVM model using all image features computed from the original mammograms, and two transformation maps increase more than 15% (from 0.79 to 0.94).

Table 3-5 Four confusion matrices generated using 4 SVMs trained using features computed from both CC and MLO view.

Feature Group	Predicted	Actual Positive	Actual Negative
SSIM	Positive	528	106
	Negative	209	1116
DCT	Positive	501	54
	Negative	1168	236
FFT	Positive	437	135
	Negative	300	1087
Fusion	Positive	**656**	61
	Negative	81	**1161**

Table 3-6 Accuracy, sensitivity, specificity, and odd ratio of using 4 SVMs trained using different features computed from both CC and MLO views.

Feature sub-group	Accuracy (%)	Sensitivity (%)	Specificity (%)	Odds Ratio
FFT features	78	59	89	11.47
DCT features	85	68	96	52
SSIM features	83	71	90	24
Fusion	**92**	**89**	**95**	**154**

Table 3-7 Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) at five subgroups with increasing values of SVM-generated prediction scores.

Subgroup (bin)	Number of Cases (Positive/Negative)	Adjusted OR	95 % CI
1	2-390	1.00	Reference
2	9-383	4.58	0.93-21.34
3	41-351	22.78	5.47-94.85
4	297-95	609.6	149.05-2493.35
5	388-3	25220	4190.90-151766.40

Table 3-8 Comparison of prediction performance of SVMs trained with and without removal of pectoral muscle area in MLO view images.

Feature sub-group	Number of features	AUC ± STD	Accuracy (%)
FFT without chest removal	14	0.70 ± 0.021	72
FFT with chest removal	14	0.84 ± 0.017	77
DCT without chest removal	14	0.68 ± 0.022	69
DCT with chest removal	14	0.89 ± 0.015	83
SSIM without chest removal	6	0.62 ± 0.026	61
SSIM with chest removal	6	0.78 ± 0.021	71
Fusion without chest removal	34	0.79 ± 0.017	83
Fusion with chest removal	34	0.94 ± 0.009	89

3.4 Discussion

This study has several unique characteristics and generates several new interesting observations. First, although many CADx schemes of mammograms (i.e., as reviewed in [8]) have been previously developed and tested to classify between malignant and benign lesions, their performance is often limited by the difficulty and errors of lesion segmentation due to the fuzziness of lesion boundary and irregular tissue overlap in 2D mammograms [10]. Since, unlike CADe schemes that aim to automatically detect suspicious lesions in which correctly cuing location of the lesion is important, CADx schemes apply to the cases in which suspicious lesions and their locations have already been detected by radiologists. The important issue in CADx schemes is to determine the likelihood of the case or the detected lesion being malignant.

However, accurately predicting the likelihood of the detected suspicious lesions being malignant remains a difficult task for radiologists, which results in higher false-positive recall rates and higher rates of benign biopsy in current clinical practice. Thus, developing a more accurate and robust CADx scheme as an assistant tool to support radiologists in their decision-making is important, no matter whether the CADx scheme uses local (region) or global image feature analysis. In this study, we explored a new approach to develop a unique case-based CADx scheme based on the detection, computation, and analysis of globally asymmetrical image features computed from two bilateral images of left and right breasts and assessed its performance using a relatively large image dataset of 1,959 cases. Thus, this new CADx scheme is a multiply image-based scheme that integrates image feature differences computed from 4 view images, which makes it significantly different from other previously single or region-based CADx schemes.

Second, we explored and tested 3 types or subgroups of global image features computed from the original FFDM images and their transformation maps, aiming to more accurately predict the likelihood of cases being malignant. From a pair of bilateral mammograms, SSIM is used in a quadratic-tree-based format that searches through different sub-blocks of original images (in a spiral way) to select areas with the highest level of bilateral asymmetry between left and right images of each case. In this way, the area outside of the breasts is removed automatically because of high SSIM values, while the area with the smallest SSIM value, which represents the highest bilateral difference, is selected. The physical meaning of the proposed SSIM-based algorithm can be well described to mimic the image features used by radiologists to assess and interpret tumors in the clinical practice. However, there is a large variation of lesion size and asymmetrical structure of the surrounding parenchymal tissue patterns in the different clinical cases. In order to automatically compensate such variations, we used an iteration approach to compute an SSIM vector with 6 SSIM index values, which represent 6 pairs of the matched sub-blocks with gradually reduced size (from 409 × 333 to 13 × 11 pixels).

The correlation coefficients of these 6 SSIM index values are relatively low (as shown in Figure 5-4). Thus, the fusion of these 6 SSIM features can increase the prediction power of using SSIM applying to a large and diverse image dataset.

Third, to further take the potential advantages of computer vision over human vision, we explored image features computed from the frequency domain. For example, FFT and DCT have been widely used as two popular frequency domains for image feature extraction in many CADx schemes to classify between malignant and benign lesions (i.e., [10]), and predict tumor response to chemotherapy (i.e., [137]). In this study, we extracted absolutely asymmetrical feature values computed from two bilateral view images (CC or MLO view of the left and right breasts) and investigated their feasibility to predict the likelihood of cases being malignant, which is also a new approach in CADx schemes of mammograms. We observed that the DCT-based features yielded the highest AUC value (as shown in Table 3-4), which shows the importance of identifying an optimal transformation map in the frequency domain for image feature extraction.

Fourth, the previous studies have reported that quantitatively detect and analysis of image features computed from MLO view images typically yielded higher performance than using image features computed using CC view images, such as applying CAD schemes to predict breast cancer risk [77] and detect suspicious lesions [138]. In this study, we systematically analyzed and compared correlation coefficients of image features computed from CC and MLO view images. The results showed that the image features computed from the MLO view images had lower distribution of the correlation coefficients than the image features computed from CC view images (Figure 3-4). This supports the results that SVMs trained using MLO view images yield higher prediction accuracy than the SVMs trained using CC view images (i.e., Table 3-3). In addition, by applying a PCA algorithm to search for and regenerate the optimal feature vectors, the best prediction performance yielded when using 65 numeric components produced by the PCA algorithm remains lower than the SVM trained using all 68 image features

computed in 3 subgroups. The results indicate that although 68 features build a relatively large feature vector or space, when considering the size of our dataset of 1,959 cases, the ratio between the numbers of the cases per class and image features remains relatively bigger (i.e., >10 per class). Thus, this size of the feature vector is acceptable in this study.

Fifth, the study also shows that image features computed from the original mammograms and transformation maps contain complementary information or discriminatory power. Thus, optimally combining or fusing multiple features computed from different feature domains to build a machine learning model or a classifier can further significantly increase CADx prediction performance (i.e., AUC value and overall accuracy after applying an operating threshold). Automatically and optimally integrating image features from the different domains is an advantage of using ML-based schemes over human observers. Additionally, the study results also show that removing pectoral muscle regions from the MLO view images can help increase prediction power to distinguish between malignant and benign cases more accurately. Thus, it is still important to develop algorithms that can more accurately and robustly detect chest walls and remove pectoral muscle regions in mammograms [131].

Sixth, many CADx studies have been previously reported in the literature to classify malignant and benign lesions. For example, reference [8] presents a table that summarizes 8 previous CADx studies, which used image datasets ranging from 38 to 1,200 cases and yielded performance of AUC values ranging from 0.70 to 0.86. Another CADx scheme that used a dataset of 560 regions of interest and a deep learning model to classify malignant and benign breast masses reported an AUC of 0.79 [12]. This study used a larger dataset involving 1,959 cases. Although we cannot directly compare the performance between this new case-based CADx scheme and previous CADx schemes reported in the literature due to the use of different image datasets, the high prediction or classification result (i.e., AUC value) of this study is encouraging. Unlike CADe schemes, which detect specific lesions and information of lesion location is important, determining lesion location is less important in CADx schemes because

the suspicious lesions have been visually detected and located by radiologists. Thus, both the conventional CADx schemes based on analysis of the image features computed from the segmented lesions and this new CADx scheme based on analysis of the global image feature difference can play the same role to support radiologists in their decision-making of predicting the likelihood of the detected lesions being malignant. By avoiding the difficulty and possible errors of breast lesion segmentation, developing a new CADx scheme based on the global mammographic image feature analysis approach can potentially be more efficient and robust.

Last, despite the encouraging results and many new observations, we recognize that this is a laboratory-based retrospective data analysis study with several limitations. First, although we assembled a relatively large and diverse image dataset, case selection bias is always an issue of concern. Second, the ratio between malignant and benign classes does not represent the actual cancer prevalence ratio in the general clinical practice. Hence, the performance and robustness of this new CADx scheme need to be further assessed and validated in future studies with new image datasets that better represent clinical practice. Third, based on the experience of our previous studies, we only explored and tested the limited numbers and types of image features, as well as the simple SVM models in this study. This may not be an optimal approach. How to identify and select optimal features and machine learning models need to be further investigated in future studies. Furthermore, this is a primary technology development study. Its clinical utility or impact on radiologists' performance in the diagnosis of breast cancer using mammograms has not been tested. In summary, despite these limitations, this study has presented a new and novel approach to develop CADx scheme based on the global image feature analysis to predict the likelihood of cases being malignant once the suspicious lesions are detected by the radiologists and demonstrated feasibility of this new approach, which may create a new opportunity for researchers in the CAD-related medical imaging informatics field to develop and optimize new computer-aided decision-making supporting tools for future clinical applications.

4 Prediction of Breast Cancer Risk Using a Machine Learning Approach Embedded with a Locality Preserving Projection Algorithm

4.1 Introduction

In spite of the heterogeneity of breast cancer, mammographic screening is widely considered the most effective approach to detect breast cancer at an early stage and help reduce the cancer mortality rate [139]. Among the existing screening methods, mammography is the only clinically accepted screening modality applied to the general population to date [140]. Despite significant advantages of mammography screening (i.e., relatively lower cost, wide accessibility, and short examination time), controversy about the population-based mammography screening remains [141] because of its lower sensitivity among a number of groups of women (i.e., women who are younger than 50 years old and have dense breasts [142]) and high false-positive recall rates [117, 143]. Thus, in order to improve efficacy of mammography screening, establishing a new personalized breast cancer screening paradigm has recently been attracting extensive research interests [144, 145]. One of the important prerequisites for realizing this goal is to identify and develop effective clinical markers or prediction tools, which have higher discriminatory power to predict the risk or likelihood of individual women having or developing image-detectable cancer in a short-term (i.e., < 2 to 5 years after a negative screening) [146].

Although a number of epidemiology-based breast cancer risk prediction models (i.e., [147, 148] models) have been developed and used to identify high-risk women, these models have low positive predictive values to help determine who should be screened in the short-term and who can be screened at longer intervals in order to increase cancer detection yield and reduce

unnecessarily frequent screening and the associated false-positive recalls with mammography [149]. Therefore, it requires identifying and developing more effective cancer risk prediction markers, including those generated from genomic tests [150] and image analysis [151, 152]. In the medical imaging field, breast density assessed from mammograms is considered an imaging marker or risk factor with much higher discriminatory power than most of the other risk factors used in the existing breast cancer risk prediction models [153]. However, subjectively rating mammographic density by radiologists based on the Breast Imaging Reporting and Data System (BIRADS) guideline is often not reliable due to large intra- and inter-reader variability [154]. In order to produce more robust results in assessing mammographic density and identifying new imaging markers to predict breast cancer risk, a number of computer-aided image processing schemes have been developed to segment and compute Volumetric Breast Density (VBD) from mammograms to predict breast cancer risk (e.g., [155]). However, whether the computed mammographic density can accurately represent breast density remains controversial [156]. The assessed mammographic density may vary due to the change of the imaging machines, imaging acquisition protocol and life cycle of the women.

In order to avoid or minimize the impact of inconsistency when using mammographic density as a breast cancer risk factor, we recently explored a new breast cancer risk factor or a quantitative imaging marker based on the bilateral asymmetry of mammographic tissue density between the left and right breasts to predict short-term breast cancer risk ([146, 157]). Since two bilateral mammograms are acquired from the same woman at one mammography screening, the relative mammographic density asymmetry is likely to remain highly consistent. From bilateral mammograms, we are able to compute a large number of image features to represent the difference in mammographic tissue density patterns. Thus, how to identify and assemble an optimal set of effective and non-redundant image features from the initial feature pool with a large number of computed image features remains a technical challenge. In order to address this challenge, the objective of this study is to develop and apply a new machine

learning approach to create a small effective feature vector for the purpose of building an optimal machine learning classifier using a relatively small training image dataset. Inspired by the deep learning technology, which directly uses input images to generate an efficient feature vector for classification [93], we proposed to apply a locally preserving projection (LPP) based feature combination algorithm [94] to reduce the dimensionality of feature space and then build a new image features based short-term breast cancer risk prediction model. Unlike the conventional feature selection methods that filter and select a set of existing optimal features from the initial feature pools, LPP generates a new optimal feature vector involving features that are different from any original features in the existing feature pool. The details of the proposed LLP approach, our image dataset, experiment and data analysis results are presented in the following sections of this article.

4.2 Materials & Method

4.2.1 Image Dataset

A testing image dataset was retrospectively assembled for this study, which includes two sets of sequential full-field digital mammography (FFDM) images acquired from 500 women participated in mammography screening. In the first set of FFDM screening, all images were determined negative by the radiologists. These negative images are named as “prior” images in this study. In the second set of FFDM screening, 250 cases were positive with cancer detected by the radiologists and verified by biopsy and histopathology tests, while the other 250 cases remained negative. All negative cases remained cancer-free for at least two more subsequent FFDM screenings. Images in the second set of screening are named as “current” images. The time interval between the “prior” and “current” mammography screenings ranged from 12 to 18 months. Although all “prior” FFDM screenings were negative, we divided 500 cases into two classes. The first class includes 250 high-risk cases in which cancer was developed and

detected in the “current” FFDM screening. The second class includes 250 low-risk cases that remained negative in the “current” FFDM screening.

Table 4-1 summarizes additional dataset information, which includes distribution of women's age and mammographic density rated by radiologists based on BIRADS guidelines. In this dataset, ages ranged between 38 and 88 years old. This is also an age-matched image dataset (≤ 1 year difference between the two classes of the cases). Thus, it has no statistically significant difference of ages between the high and low-risk case classes ($p = 0.12$). There is also no significant difference in BIRADS based mammographic density ratings between the two classes. In this study, two “prior” negative FFDM images acquired from bilateral craniocaudal (CC) view of left and right breasts were selected and used.

Table 4-1 Distribution of age and density BIRADS of cases in the dataset.

		High Risk Class	Low Risk Class
Age	Mean	58.84	57.39
	> 65 years old	57	48
	ears old	174	170
	< 45 years old	19	32
BIRADS	Extremely dense (4)	6	7
	Heterogenous (3)	133	131
	Scattered (2)	100	99
	Fatty tissue (1)	11	13

4.2.2 A Computer-aided Imaging Processing Scheme

We developed and applied a computer-aided image processing scheme to automatically segment dense fibro-glandular breast tissue regions depicted on each mammogram, and then computed bilateral mammographic tissue density and feature asymmetry between the left and right CC view images. Figure 4-1 shows a graphic user interface (GUI) of the image processing scheme. After a user selects a testing case by pointing the computer mouse to one image name

of the case and clicking the mouse button, a pair of two bilateral CC view images is uploaded into the GUI simultaneously. From each originally digital mammogram, the scheme automatically segments the breast region and generates several image maps.

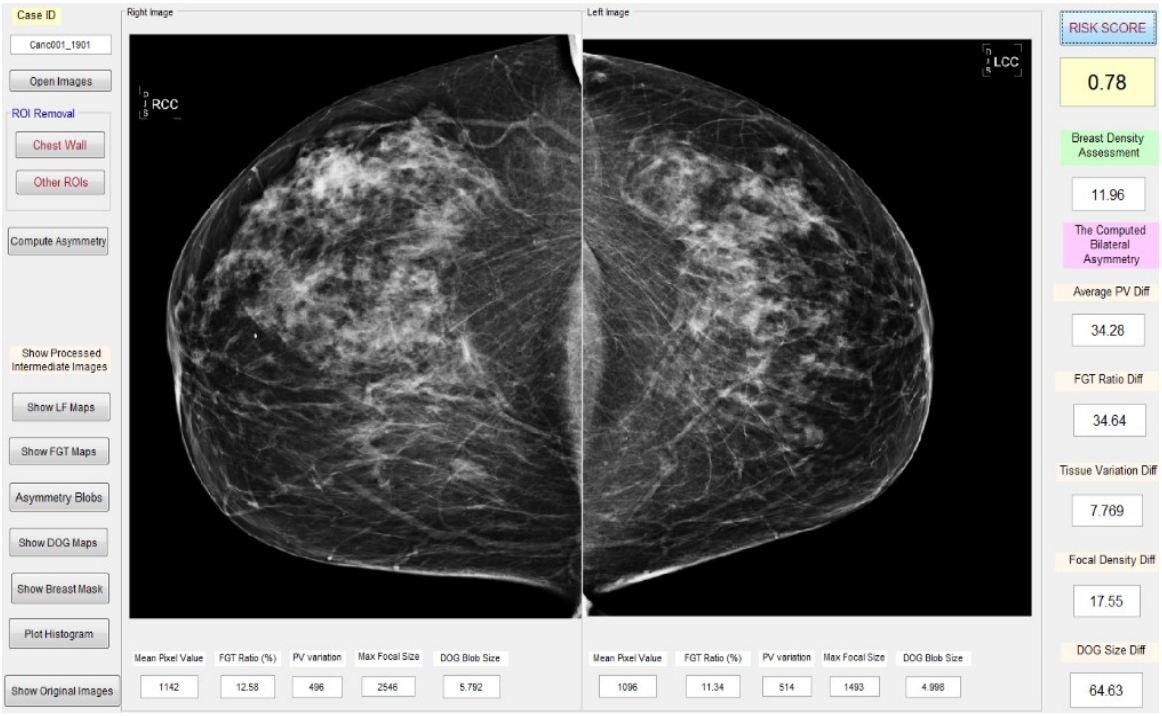


Figure 4-1 Illustration of the graphic user interface of our computer-aided imaging processing scheme to detect bilateral mammographic image feature asymmetry and predict short-term breast cancer risk.

First, a fibro-glandular tissue (FGT) density map is generated from each original mammogram as shown in Figure 4-2 (a). The percentage of FGT on each mammogram can be quantitative computed. In addition, since mammograms are two-dimensional projection images, each pixel value (or gray level) represents a percentage of the fibro-glandular tissues along the projection line (or path) of the X-ray. In order to increase the visual sensitivity to the tissue density variation, a pseudo color coding is applied to the FGT density map displayed in GUI. Second, from the FGT map, the scheme searches and segments focally dense regions as shown in Figure 4-2(b). Third, the scheme generates a local density fluctuation map as shown in Figure 4-2(c) using the method reported in our previous studies of developing computer-aided detection (CAD) scheme of mammograms [158] and mammographic image feature based

cancer risk prediction model [159]. Last, the scheme applies a difference of Gaussian (DOG) bandpass filter to generate a map showing the distribution of locally isolated small dense regions (or blobs). A similar DOG filtering map has been used in the previous CAD scheme of mammograms as the first step to detect suspicious lesions [160].

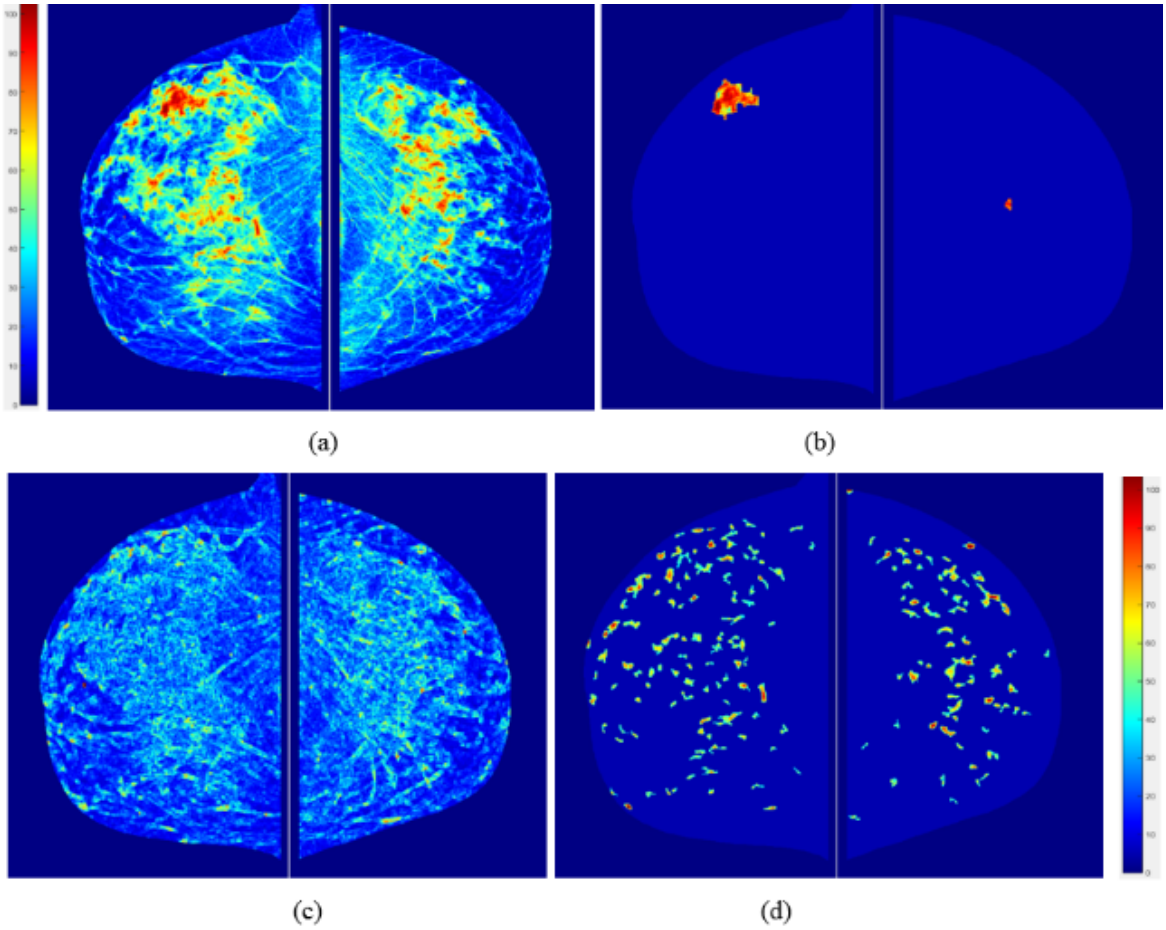


Figure 4-2 An example of showing the intermediate results of image processing steps including (a) computed breast tissue density maps, (b) detected focal density regions, (c) local density (pixel value) fluctuation maps, and (d) image maps generated using Gaussian bandpass filtering. Color bars show volumetric density level of the pixel values.

4.2.3 Image Feature Computation

Since early breast cancer usually develops in one breast, the bilateral asymmetry of mammographic tissue density or feature patterns is typically the first importantly visual sign for radiologists to detect breast abnormalities that have a high risk of leading to the cancer development. Based on the observation of how radiologists read and interpret mammograms,

we identified a new quantitative imaging marker to predict short-term breast cancer risk. Our previous study has demonstrated a trend of increasing bilateral asymmetry of the computed bilateral mammographic density features as the time lag between the negative and positive mammography screening reduces [76].

First, our computer-aided scheme calculated a series of global statistical image features related to the pixel value distribution of each image or map, which include mean, standard deviation, skewness and kurtosis of pixel value distribution in one original digital mammogram (as shown in Figure 4-1) and 4 sets of computer-processed image maps (as shown in Figure 4-2). Then, the scheme computed each bilateral asymmetry (or difference) of the image feature value by combining two corresponding feature values computed from the two bilateral images or maps between the left and right breasts. Typically, for each feature, the scheme computes and generates three combined features. First, an average feature value, $F_{ave} = \frac{1}{2} \times (F_{left} + F_{right})$, represents a global mammographic density related feature of each testing case. For example, a case with dense breasts has higher average pixel values computed from the original mammograms and FGT maps than a case of fatty breasts. Second, an absolute feature difference value, $F_{ad} = \left|F_{left} - F_{right}\right|$, or difference ratio, $F_{adr} = |F_{left} - F_{right}|/(F_{left} + F_{right})$, indicates bilateral feature asymmetry between left and right breasts. Third, a multiplication value of the above 2 features, $F_M = F_{ave} \times F_{ad}$, considers the contribution of these two factors. For example, the contribution of the bilateral asymmetry levels to the cancer risk at breasts with different mammographic density may be different. Table 4-2 lists the 44 computed image features and their definition.

Table 4-2 Description of 44 computed image features in the initial feature pool.

Image	Feature Number	Feature Description
Original FFDM image	1 – 3	Average and absolute difference of density values, and multiplication of above 2 features.
FGT map	4 - 7	Average and absolute difference of mean high density value, difference ratio and multiplication of the first 2 features.
	8 - 10	Average and absolute difference of FGT volume, and multiplication of above two features.
	11 - 13	Average and absolute difference of standard deviation of pixel values, and multiplication.
	14 - 17	Average and absolute difference of skewness and kurtosis of pixel values.
Focal density map	18 - 21	Average and absolute difference of focal density value, difference ratio and multiplication of the first 2 features.
	22 - 29	Average and absolute difference of mean, standard deviation, skewness, and kurtosis of detected and segmented focal density regions.
Local fluctuation map	29 – 33	Average of mean, standard deviation, skewness and kurtosis of pixel values.
	34 - 37	Absolute difference of mean, standard deviation, skewness and kurtosis of pixel values.
	38 - 41	Multiplication of average and absolute difference of mean, standard deviation, skewness and kurtosis of pixel values
DOG map	42 - 44	Average and absolute difference of mean pixel values, multiplication of above 2 features.

4.2.4 Machine Learning Generated Imaging Marker

Applying a machine learning method to generate an optimal and robust multi-feature fusion-based imaging marker or prediction model depends on two factors namely, (1) a set of effective and non-redundant image features, and (2) a relatively large and diverse dataset. Figure 4-3 compares 3 types of machine learning methods. First, in conventional machine learning as shown in Figure 4-3(a), segmentation and feature extraction steps are indispensable. A specific number of features from the initial feature pool are selected based on a predefined

evaluation method and index. Then, the selected features are used to build the classifier. Second, a deep learning technique does not require handcrafted features computed from the well-segmented regions. It automatically identifies features by directly learning and analyzing input images, as shown in Figure 4-3(b). However, in order to achieve robust results, the deep learning method typically requires a very large training dataset, which is often not available to provide in the cancer imaging field.

It is clear by comparing between conventional and deep machine learning methods, each has advantages and disadvantages. Conventional machine learning uses handcrafted image features, wherein it is often difficult to identify an optimal set of features that can be most effectively fused together to achieve the best performance. However, conventional machine learning is relatively easy to train using small dataset. Deep learning has the capability to automatically determine more effective features and their combination by directly learning from the input images, but its performance heavily depends on the size and diversity of the training dataset.

Figure 4-3(c) shows a new two-step approach tested in this study, which aims to take advantages of both conventional and deep machine learning approaches. Similar to a conventional learning method, this approach includes a regular region segmentation and feature extraction step to compute image features and build an initial feature pool. Then, the approach applies an LLP algorithm to learn and analyze the initially computed image features and automatically regenerate a new feature vector. This is similar to the deep learning approach, which is possible to extract new features using a deep convolution neural network (CNN) model for direct image feature learning and pass the CNN-generated features to the input layer of a conventional machine learning model, such as a support vector machine (SVM), to perform a specific classification task. However, comparing to the direct image-based deep learning technique, the number of inputs (44 features as shown in Table 4-2 as comparing to large pixel number of an input image) is significantly reduced in this study, so that a conventional machine

learning classifier embedded with the LPP image feature regeneration algorithm has potential to be more robustly trained and tested using a relatively small dataset.

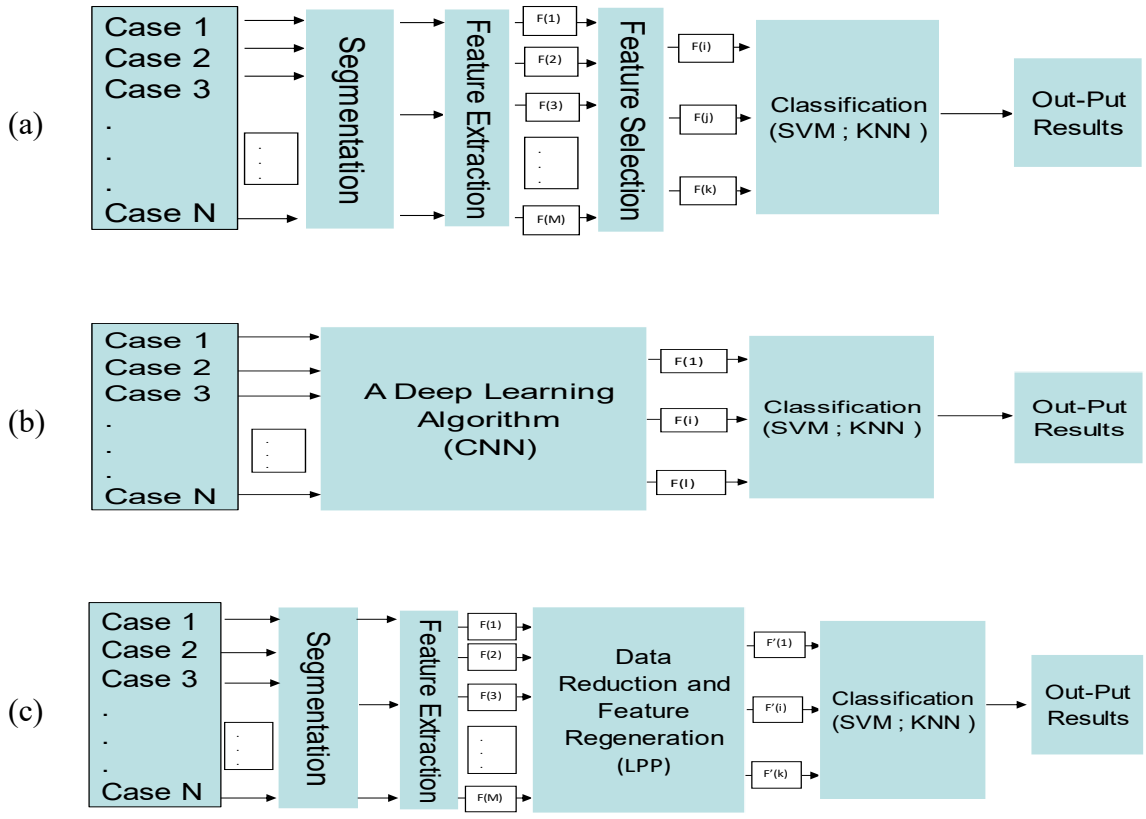


Figure 4-3 Block Diagram of three types of risk model systems, (a) conventional systems for feature selection and classification, (b) deep learning techniques for feature generation and classification, (c) proposed method for feature extraction, regeneration, and classification.

Thus, we first used a locally preserving projection (LPP) based feature combination algorithm [94] to generated new features. LPP is an unsupervised subspace learning method and a linear approximation of non-linear Laplacian eigenmaps. It involves linear projective maps to find a graph of embedding in a specific way to preserve local structure information. It has been tested and demonstrated high performance and advantages in feature dimensionality reduction, information retrieval and pattern classification or recognition. It models the manifold structure directly by constructing the nearest-neighbor graph of neighborhood information of the dataset. This graph reveals the neighborhood relations of data samples. By using the Laplacian technique, the transformation matrix of the dataset is generated to map the originally

big feature space to a much compact and more effective subspace. This linear procedure also preserves local neighborhood information of the input dataset. The design of LPP can be summarized as the following three steps [94].

1. Construct adjacent graph using ε-neighborhood or k-neighborhood.

   ε-neighborhood. The system connects nodes i and j by an edge if $\left|x_i - x_j\right|^2 < \varepsilon$.

   a) k-neighborhood: The system connects nodes i and j by an edge if i is in k NNs of j or j is in k NNs of i.

Compute weight matrix W by using either uniform weight or Gaussian weight of Euclidean distance. If nodes i and j are connected, then:

$$w_{ij} = \begin{cases} \exp\left(\left\|x_i - x_j\right\|^2 / t\right); \ x_j \in N_k(x_i), \ or \ x_i \in N_k(x_j) \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases} \qquad 4\text{-}1$$

where parameter t is a positive constant and $t \in R$, and $N_k(X_i)$ or $N_k(X_j)$ denotes a set of the k NNs of the sample x_i or x_j.

Construct the final Eigenmap. The transformation matrix P is optimized by computing the minimum eigenvalue solution to the generalized eigenvalue problem as (2):

$$XLX^T P = \lambda XDX^T P \qquad 4\text{-}2$$

where D is a diagonal matrix. A summation on the column of W makes elements of D, $D_{ii} = \sum W_{jj}$. $L = D - W$ is the Laplacian matrix. In this equation, P is composed of the optimal r projection vectors corresponding to the r smallest eigenvalues, i.e. $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_r$.

Next, we applied a conventional machine learning tool to generate a new imaging marker by optimally fusing the LPP-created new features. The learning tool was trained using our limited image dataset. In this study, we chose and compared two popular machine learning tools used in medical imaging informatics field. They are a *k*-nearest neighbor (KNN) algorithm and

a support vector machine (SVM), which use totally different learning concepts [161]. KNN is an instance-based "lazy" machine learning technique to build an optimal classification function locally. It searches for the *k* nearest training examples to classify the test sample in a pre-determined feature space and presents the class membership as the output [162]. In the KNN classifier, each test case is classified by a voting technique of its neighbors. Then, based on a distance measurement function, the case is assigned to the class most common among its *k* nearest neighbors. The Euclidean distance is used to search for the similar or nearest neighbor cases, which uses the following equation to compute distance between a queried or test case (x_q) and a selected nearest neighbor case (x_i) in a multi-feature (n) dimensional space:

$$d(y_q, x_i) = \sqrt{\sum_{r=1}^{n} \left[f_r(x_q) - f_r(x_i)\right]^2} \qquad 4\text{-}3$$

A weighting factor is defined as:

$$w_i = \frac{1}{d(x_q, x_i)^2} \qquad 4\text{-}4$$

Then, a risk prediction score or probability of the test case being a high risk (HR) case is computed as:

$$S = p(HR|x_q) = \frac{\sum_{i=1}^{K} w_i \times I(x_i = HR)}{\sum_{i=1}^{K} w_i} \qquad 4\text{-}5$$

In Equation (4-5), K indicates the number of selected nearest neighbors in the KNN prediction model and $I(x_i = HR) = 1$, when this selected nearest neighbor (x_i) is a high-risk case; otherwise, $I(x_i \neq HR) = 0$. Therefore, the risk prediction scores range from 0 (if all nearest neighbors are low-risk cases) to 1 (if all nearest neighbors are high-risk cases).

On the other hand, SVM is an "eager" machine learning method, which is trained utilizing the entire training images to build a global model of fitting the training data to predict whether a new test sample falls into one class or another [163]. Each SVM model is a representation of the data samples as points in the multi-feature space. In this space, the cases in each class are divided or separated using a hyperplane with margin of support vectors in two classes as wide as possible. Then, a new testing case is mapped to a class based on its location and distance to the hyperplane of the SVM model. Specifically, for a training data x_i ; $(i = 1, \dots, N)$, function of classifier $f(x_i)$ is introduced as (6):

$$f(x_i): \begin{cases} \geq 0 & \Rightarrow \ y_i = +1 \\ < 0 & \Rightarrow \ y_i = -1 \end{cases} \qquad \textit{4-6}$$

where y_i is the output of the system corresponding to x_i. $f(x)$ can be a linear function or other types of nonlinear functions. For instance, in linear classification, $f(x)$ can be considered as (7):

$$f(x) = \ W^T \ x + b \qquad \textit{4-7}$$

where W is weight vector and b is the bias. Among many "eager" types of machine learning classifiers (i.e., artificial neural network), SVM has advantages of building a more robust global optimization model.

In order to compute a risk prediction score of a testing case (x_q) using an SVM-based prediction model, the case is projected onto the hyperplane normal of the model. The sign distance from x_q to the decision boundary of the SVM hyperplane represents the risk prediction score of the testing case, which is computed by (8):

$$S = \sum_{j=1}^{n} a_j y_j G\left(x_j, x_q\right) + b \qquad \textit{4-8}$$

In this equation $a_j (j = 1, \dots, n), b$ are the estimated parameters of the SVM model, and $G(x_j, x_q)$ represents the dot product between x_q and the (n) support vectors (x_j). Thus, the computed risk prediction score in SVM for each testing case is its significant distance to the hyperplane. The risk prediction scores are then normalized to the range from 0 to 1 based on the maximum margin determined by the support vectors of two classes in the feature space. The higher score also indicates that the testing case is a higher risk case.

We took the following steps to combine the LPP algorithm and a machine learning classifier (either a KNN or an SVM). First, we applied the LPP algorithm to decrease the dimensionality of the feature space and rebuild the most efficient structure of features. The LLP-generated new feature vector was used as input features to build a KNN and an SVM classifier. Second, in order to reduce bias in case partitions or selection, we used a leave-one-case-out (LOCO) based cross-validation method [164] to train the classifier and test its performance. In addition, to further reduce the possible bias in the feature or data reduction and classifier training, the LPP based feature or data reduction process was embedded in the LOCO-based classifier training process to make the LPP-regenerated feature vectors independent of the testing cases. Thus, in each LOCO training and testing iteration, one case was selected from the dataset as a testing case that does not involve in the training process. LPP data reduction method was applied to the remaining training samples (i.e., 499 out of 500 samples in this study). The "best" or optimal group of features, which would be a mixture of input features to the classifier, was created by LPP to make an input feature vector for the classifier (i.e., KNN or SVM) in each training cycle. Then, the trained classifier was tested on an independent testing case by generating a risk prediction score. The higher score indicates the higher likelihood of the woman having or developing mammography-detectable breast cancer in the next subsequent mammography screening. As a result, output results are independent of input data, and results are unbiased. Similar LOCO cross-validation method with embedded feature selection or reduction has been applied and reported in our previous studies [121, 165].

4.2.5 Experiments and Performance Evaluation

In order to demonstrate potential advantages of achieving higher prediction performance using the proposed new method, we conducted a number of experiments. First, without feature selection, we trained and built the KNN and SVM-based machine learning prediction models using all 44 image features stored in the initial feature pool. Second, in order to remove low-performed image features, we computed 44 AUC values when using each of image features to predict short-term cancer risk. By sorting the computed AUC values, we selected 10 features among the top 10 AUC value list and built a new machine learning classifier. Third, we performed an exhaustive search to determine the best size of the LPP-generated feature vector. For example, in KNN, from each of K number (i.e., from 2 to 10), we systematically increased the size of the LPP-generated feature vector from 2 to 10 to search for the optimal learning parameters. Last, we tested different parameters or learning kernels used in the machine learning classifier. For example, we tested different SVMs built based on different kernel functions, including Linear, RBF, Gaussian, and Polynomial functions. Finally, the experimental results were tabulated and compared.

To evaluate the performance of the new machine learning scheme-generated risk prediction model or imaging marker, we used the following evaluation methods and indices. First, we conducted data analysis using a receiver operating characteristic (ROC) method. The area under ROC curve (AUC value) was computed and used as the evaluation index. Second, by applying an operating threshold on risk prediction scores ($T = 0.5$) to the testing data, we generated a confusion matrix with 4 parameters namely, (1) TP – true positive (high risk), (2) TN – true negative (low risk), (3) FP – false positive, and (4) FN – false-negative. From the confusion matrix, we computed overall risk prediction accuracy using the following equation.

$$Accuracy = \frac{(TP + TN)}{Total\ namber\ of\ cases} \qquad 4\text{-}9$$

Third, we sorted the risk prediction scores in an ascending order and selected 5 threshold values to divide all 500 testing cases into 5 subgroups (100 each). We computed adjusted odds ratios (ORs) and the 95% confidence intervals based on a multivariate statistical model using a publically available statistics software package (R version 2.1.1, http://www.r-project.org). An increasing trend between ORs and the classifier-generated breast cancer risk prediction scores was also computed and analyzed.

4.3 Results

When using all 44 image features included in our initial feature pool to train KNN and SVM based risk prediction classifiers, Figure 4-4 plots the distribution of cancer risk prediction accuracy when the number of neighbors (K) in the KNN classifier increases from 2 to 10, while Table 4-3 compares the difference of cancer risk prediction accuracy of applying 4 SVMs using 4 different learning kernel functions. Results showed that using K = 5 and RBF-based learning kernel yielded the highest prediction accuracy of 63.2% and 60.8% for KNN and SVM classifiers, respectively. Using these two parameters, AUC values were 0.62 and 0.60 for KNN and SVM, respectively. Thus, KNN yielded higher prediction accuracy than SVM when applying to the image dataset assembled in this study. Next, after reducing the number of input image features from 44 to 10, which are listed as top performed features, a new KNN model yielded an increased risk prediction performance with AUC = 0.64 and overall accuracy of 64.7%.

Through the exhaustive search, we identified the best parameters to develop the proposed LPP-KNN based hybrid machine learning approach (as shown in Figure 4-3c) in which (1) LPP regenerated a new feature vector with 4 features from the original feature pool of 44 features and (2) the number of neighbors in the KNN model was K = 5. Table 4-4 is a confusion matrix generated by using the optimal LPP-KNN model. Using this new prediction model, the overall cancer risk prediction accuracy further increased to 68.2%. Figure 4-5 shows the ROC curve of

using this optimal LPP-KNN model with AUC = 0.68 in comparison with the initial KNN model of using all 44 features. The increased AUC value when using this LPP-embedded machine learning approach is demonstrated.

Table 4-3 Accuracy (%) of the whole 44 feature vector for SVM classifiers with different kernel functions.

Kernel	RBF	Gaussian	Polynomial	Linear
Accuracy (%)	60.80	60.20	51.02	56.4

Table 4-4 Confusion matrix of the proposed risk model on 500 cases with Threshold = 0.5.

Actual	low-risk Cases	high-risk Cases
low-risk Cases	170	79
high-risk Cases	80	171

Table 4-5 Odds and Risk Ratio of the proposed KNN-LPP method.

Significance level	95 %
Risk Ratio	1.7597< 2.1519 <2.6316
Absolute risk reduction	36.4%
Relative risk reduction	53.5%
Odds Ratio	3.1568< 4.5997 <6.7021
Phi	0.3600
Critical Odds Ratio (COR)	1.1006

Table 4-6 Adjusted ORs and 95 % CIs for five subgroups of cases.

Number of Cases (Positive- Negative)	Adjusted OR	95 % CI
23-77	1.00	Reference
49-51	3.21	1.75-5.91
46-54	2.85	1.55-5.24
55-45	4.092	2.22-7.53
77-23	11.20	5.8-21.65

Table 4-5 summarizes several other parameters or assessment indices commonly computed and used in epidemiology studies to predict breast cancer risk. Using the threshold of (T = 0.5) to divide all testing cases into two risk classes, the odds ratio is 4.60 with a 95% confidence interval of [3.16, 6.70]. The data may indicate that women in the high-risk group have more than 4 times higher short-term breast cancer risk or probability of having or developing mammography-detectable cancer in the next subsequent breast cancer screening, which means 12 to 18 months after the "prior" negative screening of interest, than the women classified in the low-risk group.

In addition, after dividing 500 testing cases into 5 subgroups of 100 cases based on the LPP-KNN model generated cancer risk prediction scores (as shown in Table 4-6), the adjusted odd ratios increased from 1.0 in the 1st baseline subgroup of 100 cases with low-risk prediction scores to 11.2 in the 5th subgroup of 100 cases with the high-risk scores. Regression analysis result also demonstrated an increasing trend of the odds ratios with the increase in LPP-KNN model-predicted risk scores. The slope of the regression trend line between the adjusted odds ratios and the predicted risk scores is significantly different from the zero slope ($p < 0.01$).

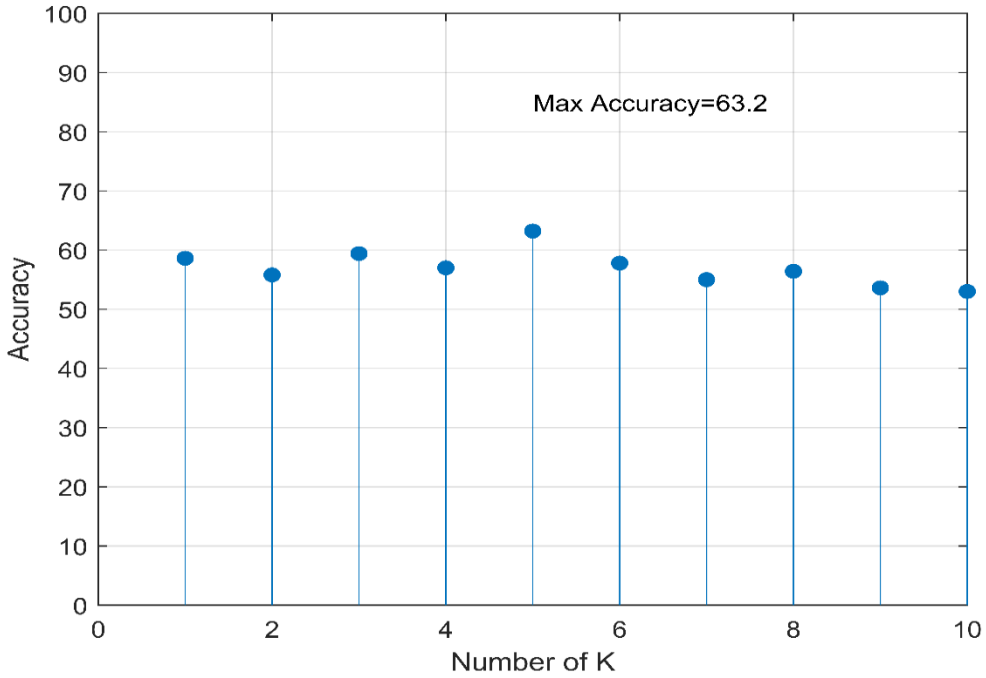


Figure 4-4 Accuracy of 44 elements feature vector with KNN classifier system for 10 different K.

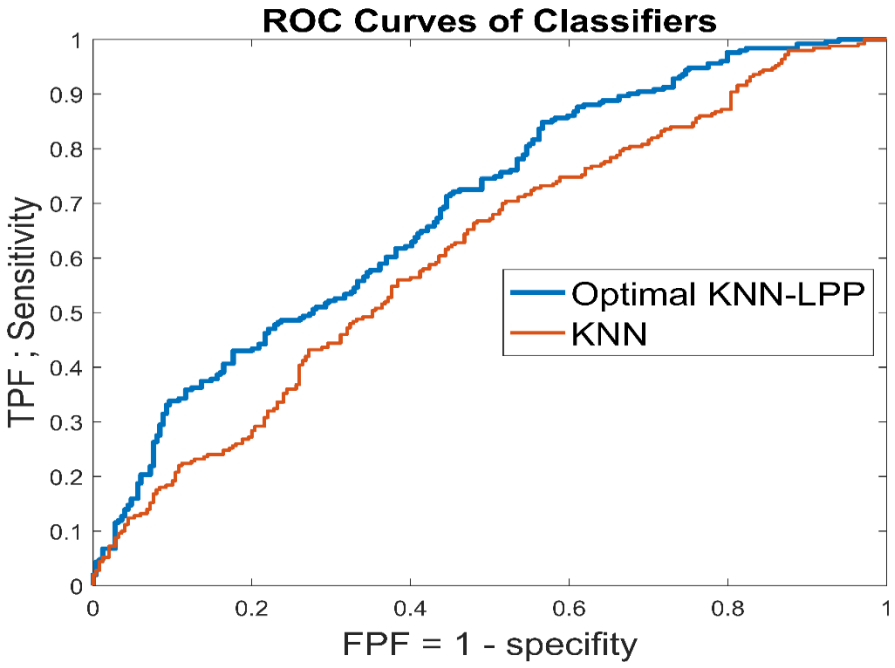


Figure 4-5 Comparison of two ROC curves generated by the original KNN model using the initial 44 features and the optimal LPP-KNN model using 4 features.

4.4 Discussion

In this study, we proposed and tested a new approach to develop a computer-aided image processing, quantitative feature analysis, and machine learning scheme for predicting short-term breast cancer risk, or the likelihood of women having or developing imaging detectable early breast cancer in the next subsequent mammography screening. This study has a number of unique characteristics compared to the previous studies reported in the literature to help improve efficacy in predicting short-term breast cancer risk and/or eventually establish a more effective personalized breast cancer screening paradigm.

First, as shown in Figure 4-3(a) and (b), two types of ML methods are commonly applied in the medical imaging informatics field to date. The main disadvantage of a conventional machine learning method is requiring many subjectively defined or "handcrafted" image features. Although deep learning (DL) can automatically define DL-generated image features by directly learning from the sample images, which may more effectively define or represent the internal structure of image data, training a robust DL model typically requires a very large image dataset. In this study, we tested a third approach, which partially takes advantage of deep learning while also maintains the advantage of the conventional machine learning to be trained using a relatively small image dataset. In our approach, as shown in Figure 4-3(c), an LPP-based feature regeneration algorithm was used to automatically learn and generate a small set of new features from a relatively large pool of initially computed image features. This process is different from the conventional feature selection, which selects optimal features from the initial feature pool (i.e., using a sequential forward floating selection (SFFS) feature selection method [136]. LPP aims to learn and redefine the effective features, which are different from any of the existing image features in the initial feature pool. Our study results demonstrated that using this LPP-based feature regeneration approach enabled us to create a smaller or compact

new feature vector and yield higher prediction performance than using either all initial image features or a set of selected highly performed features.

Second, patient age is a well-known breast cancer risk factor with the highest discriminatory power in the existing epidemiology-based breast cancer risk models [153]. Our previous studies may have bias by using the datasets in which the average age of women in the higher risk group was significantly higher than the average age in the lower risk group [146]. In order to overcome this potential bias, we in this study assembled an age-matched image dataset (as shown in Table 4-1). As a result, we removed a potential biased impact factor. The study result is encouraging by comparing to the previous studies. Specifically, although the highest adjustable odds ratio yielded in this study was very comparable or slightly higher than the results reported in our previous studies (i.e., 11.2 vs. 9.1 in [157] or 11.1 in [76]), using an age-matched dataset in this study may be important to demonstrate the robustness of developing a new optimal imaging marker based on bilateral asymmetry of mammographic tissue density between the left and right breasts.

Third, although computer-aided image processing and breast cancer risk prediction schemes had been previously developed and tested by different research groups including our own using an "eager" machine learning methods or models (i.e., artificial neural network and support vector machine), we in this study also tested a "lazy" learning method using a KNN algorithm for the purpose of predicting short-term breast cancer risk. Our study results showed that KNN can be used not only to predict cancer risk, but also to yield higher prediction accuracy that an optimized SVM model using the same testing dataset and cross-validation method. This result is quite interesting and may be worth further investigation. Using a local instance-based learning method (i.e., a KNN algorithm) can provide great flexibility to develop a new ML-based imaging marker or prediction mode because it will be relatively easy to periodically add new image data to increase size and diversity of the reference database for the

instance-based learning model, without a complicated retraining to produce a global optimization function, which is required by all other "eager" learning methods ([8, 166]).

In addition, we can also make a number of potentially interesting observations from our experimental results. For example, the highest AUC value using all 44 features was 0.62. While keeping K = 5 in the KNN learning model, removing 34 lower performed features enabled an increase of AUC value by 3.2% from 0.62 to 0.64. Furthermore, when adopting an LPP-KNN model using 4 LPP-regenerated features, AUC increased to 0.68 (representing a 9.7% increase). Thus, the results confirmed that although a large number of image features can be initially computed, removing lower performed and redundant features, as well as generating more effective features, played an important role to increase the performance of multi-feature fusion-based machine learning models. Applying LPP does not only reduce the dimension of feature space, but also it is able to reorganize the new feature vector to achieve lower amount of redundancy and maximum variance. Hence, the LPP-regenerated feature vector represents an optimal combination of the highly effective parts of all input features.

Fourth, although the size of our dataset is limited to 500 or 250 per class, applying the LPP feature regeneration approach also helped to increase the robustness of the testing result. Specifically, using the LPP approach increased the ratio between the training cases per class and image features from the original 5.7 (250/44) using all 44 features in the initial feature pool to 62.5 (250/4) using only 4 LPP-regenerated features. Thus, based on the machine learning theory, increasing this ratio will increase the robustness of the machine learning classifier to reduce the risk of overfitting. In addition, we used a leave-one-case-out (LOCO) cross-validation method to train and test the classifier, which also eliminates the bias of case partition or selection.

Despite the encouraging results, this is a proof-of-concept type study with several limitations, which need to be addressed and/or overcome in future studies. First, although LPP

is able to regenerate an optimal image feature vector, its ultimate performance depends on the quality of initial feature pool. The initial feature pool with 44 features used in this study may not have been an optimal feature pool. Thus, we will continue our efforts to improve the computer-aided image processing scheme to more accurately and robustly segment dense mammographic tissue regions and compute image features. Second, due to the potential compression difference between left and right breasts, breast sizes and density overlapping ratio (or pixel values) depicting on two bilateral images may not be the same. In order to reduce the potential errors in computing feature difference, we need to continue investigating new methods to compensate the difference and reduce the errors. Third, since the regions near breast skin and behind chest wall have high pixel values in mammograms, in order to avoid adding them into the dense fibro-glandular tissue volume, we need to develop a more accurate method to automatically remove these regions without losing significant information of breast area. Fourth, it is also important to more effectively detect and compensate other types of image noise, which may exist and vary in screening mammograms due to the variety of technical issues in conducting mammography examinations on different individual women. The goal is to develop a more robust computer-aided image processing scheme to achieve high accuracy in mammographic dense tissue segmentation. Last, this study only used and analyzed images acquired from one "prior" mammography screening. In the future studies, we will collect more cases with multiple "prior" mammography screenings and investigate the feasibility of improving the performance of short-term breast cancer risk prediction by combining the image feature variation trend among the multiple mammography screenings into the risk prediction models.

5 Applying a Random Projection Algorithm to Optimize Machine Learning Model for Breast Lesion Classification

5.1 Introduction

Developing Computer-aided detection and diagnosis (CAD) schemes of medical images have been attracting broad research interest in order to detect suspicious diseased regions, classify between malignant and benign lesions, quantify disease severity, and predict disease prognosis or monitor treatment efficacy. Some CAD schemes have been used as "a second reader" or quantitative image marker assessment tools in clinical practice to assist clinicians (i.e., radiologists) aiming to improve image reading accuracy and reduce the inter-reader variability [167]. Despite of extensive research effort and progress made in the CAD field, researchers still face many challenges in developing CAD schemes for clinical applications [1, 168, 169]. For example, in developing CAD schemes, machine learning plays a critical role, which use image features to train classification models to predict the likelihood of the analyzed regions depicting or patterns representing diseases. However, due to the great heterogeneity of disease patterns and the limited size of image datasets, how to identify a small and optimal image feature vector to build the highly performed and robust machine learning models remains a difficult task.

In current CAD schemes, after image preprocessing to reduce image noise, detecting and segmenting suspicious regions of interest (ROIs), CAD schemes can compute many image features from the entire image region or the segmented ROIs [170]. Recently, two methods have attracted broad research interest to compute image features. One uses a deep transfer learning model as an automated feature extractor (i.e., extracting 4,096 features in a fully connected layer (FC6 or FC7) of an AlexNet). The disadvantage of this approach is requiring very big training and validation image datasets, which are often not available in medical image fields.

Another approach uses radiomics concept and method to compute and generate an initial feature pool. Although Radiomics typically computes smaller number of features than deep learning based feature extractors, it may still compute many features (i.e., >1,000 image features, which mostly represent texture patterns of the segmented ROIs in variety of scanning directions as reported in previous studies [171, 172]). However, due to the limited size of the training datasets, such large number of image features can often drive to overfit machine learning models and reduce model robustness. Thus, it is important to build an optimal feature vector from the initially large feature pool in which the generated features should not be redundant or highly correlated [37, 173]. Then, machine learning models can be better trained to achieve the enhanced performance and robustness. In general, if the feature dimensionality reduction happens with choosing the most effective image features from the initial feature pool, it is known as feature selection (i.e., using sequential forward floating selection (SFFS) [136]). On the other hand, if the dimensionality reduction comes from reanalyzing the initial set of features to produce a new set of orthogonal features, it is known as feature regeneration (i.e., principal component analysis (PCA) and its modified algorithms [95]). Comparing between these two methods, feature regeneration method has advantages to more effectively eliminate or reduce redundancy or correlation in the final optimal image feature vector. However, most of medical image data or features have very complicated or heterogeneous distribution patterns, which may not meet the precondition that all feature variables are linear to optimally apply PCA-type feature regeneration methods.

In order to better address this challenge and more reliably regenerate image feature vector for developing CAD schemes of medical images, we investigate and test another feature regeneration method namely, a random projection algorithm (RPA), which is an efficient way to map features into a space with a lower-dimensional subspace, while preserving the distances between points under better contrast. This mapping process is done with a random projection matrix. In the lower space since the distance is preserved, it will be much easier and reliably to

classify between two feature classes. Because of its advantages and high performance, RPA has been tested and implemented in a wide range of engineering applications including handwrite recognition [97], face recognition and detection [98], visual object tracking and recognition [99, 100], and car detection [101].

Thus, motivated by the success of applying RPA to the complex and nonlinear feature data used in many engineering application domains, we hypothesize that RPA also has advantages when applying to medical images with the heterogeneous feature distributions. To test our hypothesis, we conduct this study to investigate feasibility and potential advantages of applying RPA to build optimal feature vector and train machine learning model implemented in a new computer-aided diagnosis (CAD) scheme to classify between malignant and benign breast lesions depicting on digital mammograms. The details of the assembled image dataset, the experimental methods of feature regeneration using RPA and a support vector machine (SVM) model optimization, data analysis and performance evaluation results are presented in the following sections.

5.2 Materials and Methods

5.2.1 Image Dataset

A fully anonymized dataset of full-field digital mammography (FFDM) images acquired from 1,487 patients are retrospectively assembled and used in this study. All cases were randomly selected by an institutional review board (IRB) certified research coordinator from the cancer repository and picture archive and communication system (PACS). All selected cases have suspicious soft-tissue mass type lesions previously detected by the radiologists on the mammograms. Based on lesion biopsy results, 644 cases depict malignant lesions and 843 cases had benign lesions. These patients have an age range from 35 to 80 years old. Table 5-1 summarizes and compares case distribution information of patients' age and mammographic density rated by radiologists using breast imaging reporting and data system (BIRADS)

guidelines. As shown in the table, patients in benign group are moderately younger than the patients in the malignant group. However, there is not a significant difference of mammographic density between the two groups of patients (p = 0.576).

Table 5-1 Case numbers and percentage distribution of patients' age and mammographic density rated by radiologists using BIRADS guidelines.

	Subgroup	Malignant Cases	Benign Cases
Density BIRADS			
	1	25 (3.9%)	58 (6.9%)
	2	186 (28.8%)	262 (31.1%)
	3	401 (62.3%)	502 (59.5%)
p-value = 0.576	4	32 (5.0%)	21 (2.5%)
Age of Patients (years old)			
	A < 40	11 (3.4%)	71 (8.4%)
	$40 \leq A < 50$	109(19.2%)	158(18.7%)
	$50 \leq A < 60$	167(25.6%)	285(33.8%)
	$60 \leq A < 70$	180(24.4%)	192(22.8%)
	$70 \leq A$	177(27.4%)	137(16.3%)

All FFDM images were acquired using one type of digital mammography machines (Selenia Dimensions made by the Hologic Company), which have a fixed pixel size of $70\mu m$ in order to detect microcalcifications. Since in this study, we only focus on classification of soft tissue mass type lesions, all images are thus subsampled using a pixel averaging method with a 5×5 pixel frame, so that the pixel size of the subsampled images increases to 0.35mm. This subsample method has been used and reported in many of our previous CAD studies (i.e., [83, 124]). Additionally, in this dataset, the majority of cases have two craniocaudal (CC) and mediolateral oblique (MLO) view mammograms of either left or right breast in which the suspicious lesions are detected by the radiologists, while small fraction of cases just have one CC or MLO image in which the lesions were detected. Overall, 1,197 images depicting malignant lesions and 1,302 images depicting benign lesions are available in this image dataset. All lesion centers are visually marked by the radiologists using a custom-designed interactive graphic user interface (GUI) tool. The marked lesion centers are recorded and used as "ground-truth" to evaluate CAD performance [124].

5.2.2 Initial Image Feature Pool with a High Dimensionality

In developing CAD schemes to classify between malignant and benign breast lesions, many different approaches have been investigated and applied to compute image features including those computed from the segmented lesions [10], the fixed regions of interest (ROIs) [174] and the entire breast area [83]. Each approach has advantages and disadvantages. However, their classification performance may be quite comparable with an appropriate training and optimization process. Thus, since this study focus on investigating the feasibility and potential advantages of a new feature dimensionality reduction method of RPA, we will use a simple approach to compute the initial image features from both the fixed ROI and the segmented lesion regions.

Since classification between malignant and benign lesions is a difficult task, which depends on optimal fusion of many image features related to tissue density heterogeneity, speculation of lesion boundary, as well as variation of surrounding tissues. Previous studies have demonstrated that statistics and texture features can be used to model these valuable image features including intensity, energy, uniformity, entropy, and statistical moments, etc. Thus, like most CAD schemes using the ROIs with a fixed size as classification targets (including the schemes using deep learning approaches [12]), this CAD scheme also focuses on using the statistics and texture-based image features computed from the defined ROIs and the segmented lesion regions. For this purpose, following methods are used to compute image features that are included in the initial feature pool.

First, from a ROI of an input image, gray level difference method (GLDM) is used to compute the occurrence of the absolute difference between pairs of gray levels divided in a particularly defined distance in several directions. It is a practical way for modeling analytical texture features. The output of this function is four different probability distributions. For an image $I(m,n)$, we consider displacement in different directions like $\delta(d_x, d_y)$, then $\hat{I}(m,n) =$

$\left|I(m,n) - I(m+d_x, n+d_y)\right|$ estimates the absolute difference between gray levels, where d_x, d_y are integer values. Now it is possible to determine an estimated probability density function for $\hat{I}(m,n)$ like $f(.|\delta)$ in which $f(i|\delta) = P(\hat{I}(m,n) = i)$. It means for an image with L gray levels, the probability density function is L-dimensional. The components in each index of the function show the probability of $\hat{I}(m,n)$ with the same value of the index. In the proposed method implemented in this CAD study, we consider $d_x = d_y = 11$, which is calculated heuristically [175]. The probability functions are computed in four directions ($\varphi = 0, \pi/4, \pi/2, 3\pi/4$), which signifies that four probability functions are computed to provide the absolute differences in four primary directions that each of which is used for feature extraction.

Second, a gray-level co-occurrence matrix (GLCM) estimates the second-order joint conditional probability density function. The GLCM carries information about the locations of pixels having similar gray level values, as well as the distance and angular spatial correlation over an image sub-region. To establish the occurrence probability of pixels with the gray level of i, j over an image along a given distance of d and a specific orientation of φ, we have $P(i,j,d,\varphi)$. In this way, the output matrix has a dimension of the gray levels (L) of the image [176]. Like GLDM, we compute four co-occurrence matrices in four cardinal directions ($\varphi = 0,\ \pi/4,\ \pi/2,\ 3\pi/4$). GLCM is rotation invariant. We combine the results of different angles in a summation mode to obtain the following probability density function for feature extraction, which is also normalized to reduce image dependence.

$$P(i,j) = \sum_{\varphi=0,\pi/4,\pi/2,3\pi/4} P(i,j,d=2,\varphi)$$
$$P(i,j) = \frac{P(i,j)}{\sum_i \sum_j P(i,j)}; i,j = 1,2,3,\dots,L \qquad \textit{5-1}$$

Third, a gray level run length matrix (GLRLM) is another popular way to extract textural features. In each local area depicting suspicious breast lesion, a set of pixel values are searched

within a predefined interval of the gray levels in several directions. They are defined as gray level runs. GLRM calculates the length of gray-level runs. The length of the run is the number of pixels within the run. In the ROI, spatial variation of the pixel values for benign and malignant lesions may be different, and gray level run is a proper way to delineate this variation. The output of a GLRM is a matrix with elements that express the number of runs in a particular gray level interval with a distinct length. Depending on the orientation of the run, different matrices can be formed [177]. We in this study consider four different directions ($\varphi = 0, \pi/4, \pi/2, 3\pi/4$) for GLRM calculations. Then, just like GLCM, GLRM is also rotation invariant. Thus, the output matrices of different angles in a summation mode are merged to generate one matrix.

Fourth, in addition to the computing texture features from the ROI of the original image in the spatial domain, we also explore and conduct multiresolution analysis, which is a reliable way to make it possible to perform zooming concept through a wide range of sub-bands in more details [178]. Hence, textural features extracted from the multiresolution sub-bands manifest the difference in texture more clearly. Specifically, a wavelet transform is performed to extract image texture features. Wavelet decomposes an image into the sub-bands made with high-pass and low-pass filters in horizontal and vertical directions followed by a down-sampling process. While down-sampling is suitable for noise cancelation and data compression, high-pass filters are beneficial to focus on edge, variations, and the deviation, which can show and quantify texture difference between benign and malignant lesions. For this purpose, we apply 2D Daubechies (Db4) wavelet on each ROI to get approximate and detailed coefficients. From the computed wavelet maps, a wide range of texture features is extracted from principal components of this domain.

Moreover, analyzing geometry and boundary of the breast lesions and the neighboring area is another way to distinguish benign and malignant lesions. In general, benign lesions are typically round, smooth, convex shaped, with well-circumscribed boundary, while malignant

lesions tend to be much blurry, irregular, rough, with non-convex shapes [179]. Hence, we also extract and compute a group of features that represent geometry and shape of lesion boundary contour. Then, we add all computed features as described above to create the initial pool of image features.

5.2.3 Applying Random Projection Algorithm (RPA) to Generate Optimal Feature Vector

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 two points like $X = (x_1, \dots, x_d)$, and $Y = (y_1, \dots y_d)$, the distance in d dimensional spaces define as:

$$|X - Y| = \sqrt{\sum_{j=1}^{d} (x_j - y_j)^2} \qquad 5\text{-}2$$

In addition, it is also possible to define the volume V of a sphere in a d dimensional space as a function of its radius (r) and the dimension of the space as (3). This equation is proved in [180].

$$V(d) = \frac{r^d \pi^{\frac{d}{2}}}{\frac{1}{2} d\Gamma(\frac{d}{2})} \qquad 5\text{-}3$$

The matrix of features is normalized between [0, 1]. It means a sphere with $r = 1$ can encompass all the data. An interesting fact about a unit-radius sphere is that as equation (5-4) shows, as the dimension increase, the volume goes to zero. Since $\pi^{\frac{d}{2}}$ is an exponential of $\frac{d}{2}$, while growing rate of $\Gamma(\frac{d}{2})$ is a factorial of $\frac{d}{2}$. At the same time, the maximum possible distance between two points stays at 2.

$$\lim_{d\to\infty}\left(\frac{\pi^{\frac{d}{2}}}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)}\right)\cong 0 \qquad 5\text{-}4$$

Moreover, based on the heavy-tailed distribution theorem, for a case like $X=(x_1,\dots,x_d)$ in the space of features, suppose with an acceptable approximation features are independent, or nearly perpendicular variables as mapped to different axes, with $E(x_i)=p_i$, $\sum_{i=1}^{d}p_i=\mu$ and $E|(x_i-p_i)^k|\le p_i$ for $k=2,3,\dots,\lfloor t^2/6\mu\rfloor$, then, the previous study [181] has proven that:

$$prob\left(\left|\sum_{i=1}^{d}x_i-\mu\right|\ge t\right)\le Max\left(3e^{\frac{-t^2}{12\mu}},4\times 2^{\frac{-t}{e}}\right) \qquad 5\text{-}5$$

We can perceive that the farther the value of t increases, the smaller the chance of having a point out of that distance, which means that *X* would be concentrated around the mean value. Overall, based on equations (5-4), and (5-5) with an acceptable approximation, all data are encompassed in a sphere of size one, and they are concentrated around their mean value. As a result, if the dimensionality is high, the volume of the sphere is close to zero. Hence, the contrast between the cases is not enough for a proper classification.

Above analysis also indicates the more features included in the initial feature vector, the higher the dimension of the space is, and the more data is concentrated around the center, which makes it more difficult to have enough contrast between the features. A powerful technique to reduce the dimensionality while approximately preserves the distance between the points, which implies approximate preservation of the highest amount of information, is the key point that we are looking for. If we adopt a typical feature selection method and randomly select a k-dimensional sup-space of the initial feature vector, it is possible to prove that all the projected distances in the new space are within a determined scale-factor of the initial d-dimensional space [182]. Hence, although some redundant features are removed, the final accuracy may not increase, since contrast between the points may still be not enough to present a robust model.

To address this issue, we take advantage of Johnson-Lindenstrauss Lemma to optimize the feature space. Based on the idea of this lemma, for any $0 < \epsilon < 1$, and any number of cases as N, which are like the points in d-dimensional space ($\mathbb{R}^d$), if we assume k as a positive integer, it can be computed as:

$$k \geq 4\frac{\ln N}{(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})} \qquad 5\text{-}6$$

Then, for any set V of N points in $\mathbb{R}^d$, for all $u, v \in V$, it is possible to prove that there is a map, or random projection function like $f: \mathbb{R}^d \to \mathbb{R}^k$, which preserves the distance in the following approximation [183], which is known as Restricted Isometry Property(RIP):

$$(1 - \epsilon)|u - v|^2 \leq |f(u) - f(v)|^2 \leq (1 + \epsilon)|u - v|^2 \qquad 5\text{-}7$$

Another arrangement of this formula is like:

$$\frac{|f(u) - f(v)|^2}{(1 + \epsilon)} \leq |u - v|^2 \leq \frac{|f(u) - f(v)|^2}{(1 - \epsilon)} \qquad 5\text{-}8$$

As these formulas show the distance between the set of points in the lower-dimension space is approximately close to the distance in high-dimensional space. This Lemma states that it is possible to project a set of points from a high-dimensional space into a lower dimensional space, while the distances between the points are nearly preserved.

It implies that if we project the initial group of features into a space with a lower-dimensional subspace using the random projection method, the distances between points are preserved under better contrast. This may help better classify between two feature classes representing benign and malignant lesions with low risk of overfitting.

It should be noted that for an input matrix of features like $X \in \mathbb{R}^{n \times d}$, n and d represent the number of training samples and features, respectively. Unlike the principal component analysis (PCA) that assumes relationship among feature variables are linear and intends to generate new orthogonal features, RPA aims to preserve distance of the points (training samples) while

reducing the space dimensionality. Thus, using RPA will create a subspace $\tilde{X} = XR$ in which R satisfies the RIP condition, and $R \in \mathbb{R}^{d\times k}, \tilde{X} \in \mathbb{R}^{n\times k}$. Since the subspace's geometry is preserved, previous studies [184, 185] proved that a SVM based machine learning classifier could better preserve the characteristics of the image dataset to build the optimal hyperplane and thus reduce the generalization error. In other words, if an SVM classifier makes the resulting margin $\gamma^* = {}^{1}/_{\|w^*\|^2}$ for its optimal hyperplane (w^*) after solving the optimization problem on the initial feature space of X, and on the subspace of $\tilde{X}$, it makes the resulting margin $\tilde{\gamma}^* = {}^{1}/_{\|\tilde{w}^*\|^2}$ for the respective optimized hyperplane ($\tilde{w}^*$). Another study [186] proved that hinge loss (for margin $\tilde{\gamma}^*$) of the classifier trained on the subspace data ($\tilde{X}$) is less than that (γ^*) of the classifier trained on the original data (X). Strictly speaking, the trained classifier's error rate on the optimized subspace generated using RPA is lower than that of the classifier trained on the original space. It indicates that training a machine learning classifier using an optimal subspace under RIP condition can build a more accurate and robust model for the classification purpose.

In this study, we investigate and demonstrate whether using RPA can yield better results as compared to other popular feature dimensionality reduction approaches (i.e., PCA).

5.2.4 Experiment of Feature Combination and Dimensionality Reduction

First, the proposed CAD scheme applies an image preprocessing step to the whole images in the dataset to read them one by one, and based on the lesion centers pre-marked by the radiologists to extract a squared ROI area in which the centers of the lesion and ROI overlap. In order to identify the optimal size of the ROIs, a heuristic method is applied to select and analyze ROI size. Basically, the different ROI sizes (i.e., in the range from 128×128 to 180×180 pixels) are examined and compared. From the experiments, we observe that the ROIs with size of 150×150 pixels generate the best classification results applying to this large and diverse

dataset, which reveals that this is the most efficient size to cover all mass lesions included in our diverse dataset, which corresponds to use the ROI of $52.5 \times 52.5mm^2$. Figure 5-1 shows examples of 4 ROIs depicting two malignant lesions and two benign lesions. After ROI determination, all the images in the dataset are saved in Portable Network Graphics (PNG) format with 16 bits in the lossless mode for the feature extraction phase.

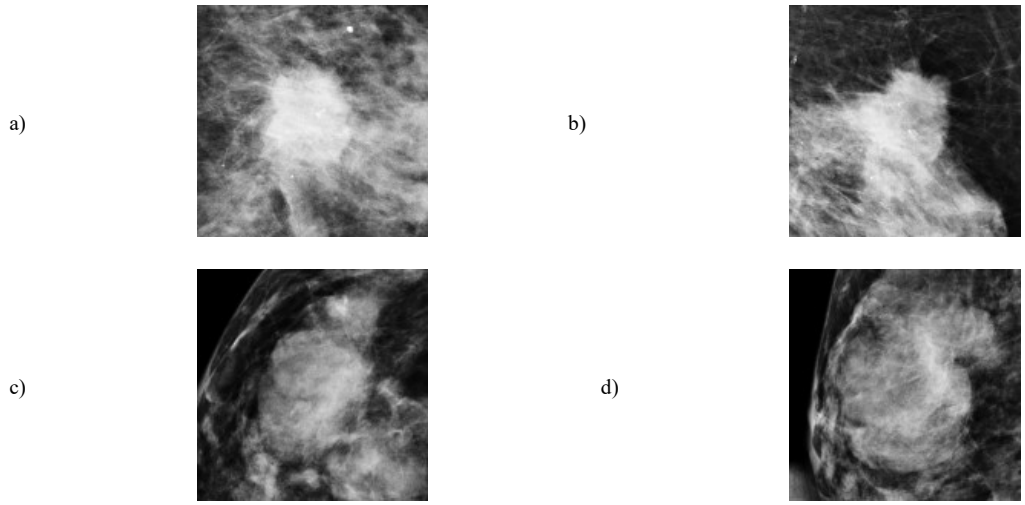


Figure 5-1 Example of 4 extracted ROIs with the detected suspicious soft-tissue masses (lesions) in ROI center. a,b) 2 ROIs involving malignant lesions and c,d) 2 ROIs involving benign lesions.

Next, the CAD scheme is applied to segment lesion from the background. For this process, CAD applies an unsharp masking method in which a low-pass filter with a window-size of 30 is first applied to filter the whole ROI. Next, CAD computes the absolute pixel value difference between the original ROI and the filtered ROI to produce a new image map that highlights the lesion and other regions (or blobs) with locally higher and heterogeneous tissue density. Then, CAD applies morphological filters (i.e., opening and closing) to delete the small and isolated blobs (with the pixel members less than 50), and repair boundary contour of the lesion and other remaining blobs with higher tissue density. Since in this study, the user clicks the lesion center and the ROI is extracted around this clicked point, the blob located in the center of ROI represents the segmented lesion. Figure 5-2 shows an example of applying this algorithm to locate and segment suspicious lesion from the surrounding tissue background.

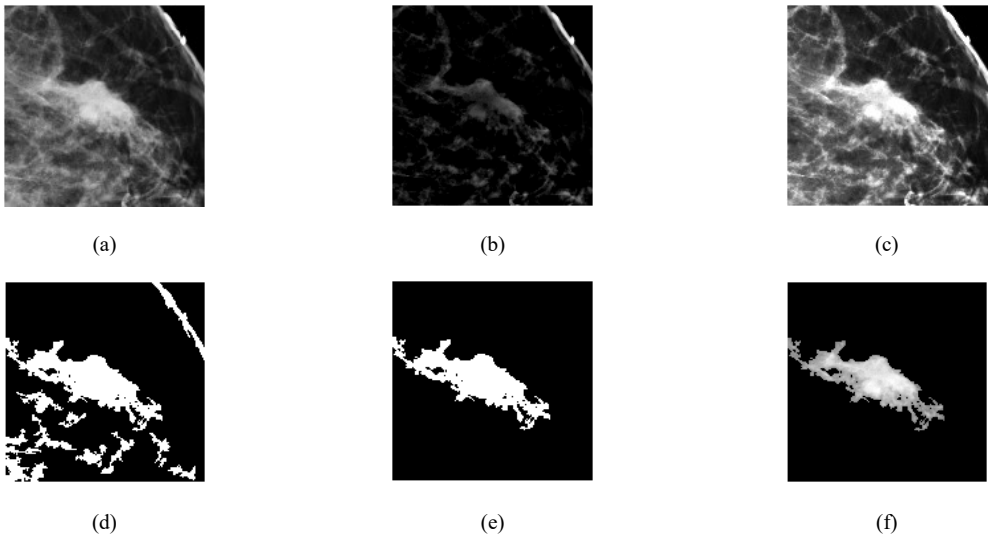


Figure 5-2 Example to illustrate lesion segmentation, which include a) the original ROI, b) absolute difference of ROI from low-pass filtered version, c) combination of (a) and (b) which gives the suspicious regions better contrast to the background, d) output of of morphological filtering, e) blob with the largest size is selected (a binary version of the lesion), and f) finally segmented lesion area. It is output of mapping (e) to (a).

Table 5-2 List of computed features on ROI area

Feature category	Feature Description
Density related features	1.Mean, 2. variance, 3. skewness, 4. kurtosis, 5. entropy, 6. correlation, 7. energy, 8. root mean square level, 9. uniformity, 10. max, 11. min, 12. median, 13. range, 14. mean absolute deviation, 15. Contrast, 16. homogeneity, 17. smoothness, 18. inverse difference movement, 19. suspicious regions volume, 20. standard deviation.

After image segmentation, CAD scheme computes several sets of the relevant image features. The first group of features are the pixel value (or density) related statistics features as summarized in Table 5-2. These 20 statistics features are repeatedly computed from three types of images namely, 1) the entire ROI of the original images (as shown in Figure 5-2(a)), 2) the segmented lesion region (as shown in Figure 5-2(f)), and 3) all highly dense and heterogeneous tissue blobs (as shown in Figure 5-2(d)). Thus, this group of features includes 60 statistics features.

The second group of features is computed from the GLRLM matrix of the ROI area. For this purpose, 16 different quantization levels are considered to calculate all probability functions in four different directions from the histograms. After combining the probability

functions, on rotation invariance version of them, the following group of features is computed. Features are short-run emphasis, long-run emphasis, gray level non-uniformity, run percentage, run-length non-uniformity, low gray level run emphasis, and high gray level run emphasis. Hence, this group of features includes seven GLRM-based features.

The third group of features includes GLDM based features computed from the entire ROI. Specifically, we select a distance value of 11 pixels for the inter-sample distance calculation. CAD computes four different probability density functions (PDFs) based on the image histogram calculation in different directions. The PDF (p) with (μ) as the mean of the population, standard deviation, root mean square level, and the first four statistical moments ($n = 1, 2, 3, 4$) with the following equation are calculated as features.

$$\widehat{m}_n = \sum_{i=1}^{N} p_i(x_i - \mu)^n \qquad \textit{5-9}$$

It is an unbiased estimate of nth moment possible to calculate by:

$$m_n = \int_{-\infty}^{\infty} p(x)x^n dx \qquad \textit{5-10}$$

As shown in equation 10, $p(x)$ is weighted by x^n. Hence, any change in the *p(x)* is polynomially reinforced in the statistical moments. Thus, any difference in the four PDFs computed from malignant lesions is likely to be polynomially reinforced in the statistical moments of the computed coefficients. Six features from each of four GLDM based PDFs make this feature group, which has total 24 features.

The fourth group of features computes GLCM based texture feature. Based on the method proposed in the previous study [187], our CAD scheme generates a matrix of 44 textural features computed from GLCM matrix based on all GLCM based equations proposed in [176].

In this way any properties of the GLCM matrix proper for the classification purpose is granted. Hence, this group contains 44 features computed from the entire ROI.

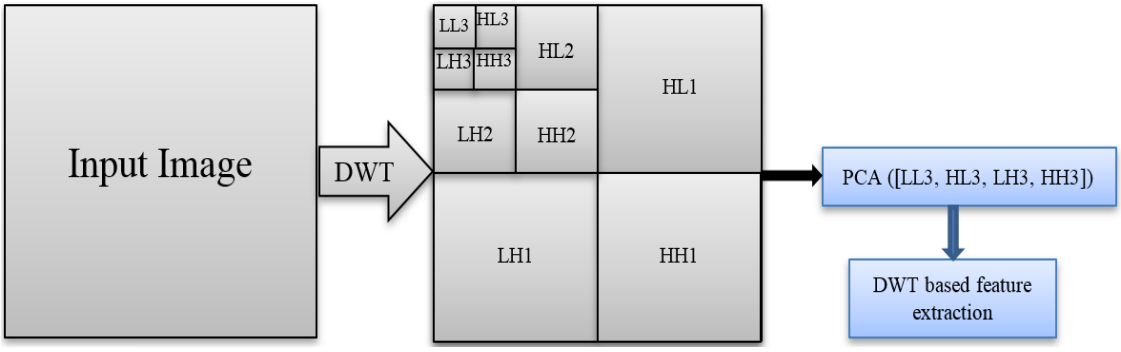


Figure 5-3 Wavelet based feature extraction. Wavelet decomposition is applied three times to make the images compress as possible. Then PCA is adopted as another way of data compression.

Table 5-3 List of Wavelet-based extracted features.

Feature category	Feature Description
Wavelet-based features	1. Contrast, 2. Correlation, 3. Energy, 4. Homogeneity, 5. Mean, 6. Standard deviation, 7. Entropy, 8. Root mean square level, 9. Variance, 10. Smoothness, 11. Kurtosis, 12. Skewness, 13. IDM

Table 5-4 List of Geometrical-based extracted features.

Feature category	Feature Description
Geometrical based features	1. Area, 2. Major Axis Length, 3. Minor Axis Length, 4. Eccentricity, 5. Orientation, 6. Convex Area, 7. Circularity, 8. Filled Area, 9. Euler Number, 10. Equivalent Diameter, 11. Solidity, 12. Extent, 13. Perimeter, 14. Perimeter Old,15. Max Feret Diameter,16. Max Feret Angle,18. Min Feret Diameter,19. Min Feret Angle, 20. Roundness Ratio.

The fifth group of features includes wavelet-based features. The Daubechies wavelet decomposition is accomplished on the original ROI (i.e., Figure 5-2(a)). Figure 5-3 shows a block diagram of the wavelet-based feature extraction procedure. The last four sub-bands of wavelet transform are used to build a matrix of four sub-bands in which principal components of this matrix are driven for feature extraction and computation. The computed features are listed in Table 5-3. We also repeat the same process to compute wavelet-based feature from the

segmented lesion (i.e., Figure 5-2(f)). As a result, this feature group includes 26 wavelet-based image features.

Last, to address the differences between morphological and structural characteristics of benign and malignant lesions, another group of geometrical based features is derived and computed from the segmented lesion region. For this purpose, a binary version of the lesion, like what we showed in Figure 5-2 (e), is first segmented from the ROI area. Then, all the properties listed in Table 5-4 are calculated from the segmented lesion region in the image using the equations reported in [188].

By combining all features computed in above 6 groups, CAD scheme creates an initial pool of 181 image features. Then, RPA is applied to reduce feature dimensionality and generate an optimal feature vector. For this purpose, we utilize sparse random matrix as the projection function to achieve the criteria as defined in equation (5-7). Sparse random matrix is a memory efficient and fast computing way of projecting data, which guarantees the embedding quality of this idea. To do so, if we define $s = 1/density$, in which $density$ defines ratio of non-zero components in the RPA, the components of the matrix as random matrix elements (RME) are:

$$RME = \begin{cases} -\sqrt{\frac{s}{n_{components}}}, & {}^{1}/_{2s} \\ 0, \quad with\ probability & 1 - {}^{1}/_{s} \\ \sqrt{\frac{s}{n_{components}}}, & {}^{1}/_{2s} \end{cases} \qquad 5\text{-}11$$

In this process, we select $n_{components}$, which is the size of the projected subspace. As recommended in [189], we consider number of non-zero elements to the minimum density, which is: ${}^{1}/\sqrt{n_features}$.

5.2.5 Development and Evaluation of Machine Learning Model

After processing images and computing image features from all 1,197 ROIs depicting malignant lesions and 1,302 ROIs depicting benign lesions, we build machine learning model to classify between malignant and benign lesions by taking following steps or measures. Figure 5-4 shows a block diagram of the machine learning model along with the training

and testing process. First, although many machine learning models (i.e., artificial neural networks, K-nearest neighborhood network, Bayesian belief network, support vector machine) have been investigated and used to develop CAD schemes, based on our previous research experience [83], we adopt the support vector machine (SVM) to train a multi-feature fusion based machine leaning model to predict the likelihood of lesions being malignancy in this study. Under a grid search and hyperparameter analyses, linear kernel implemented in SVM model can achieve a low computational cost and high robustness in prediction results as well.

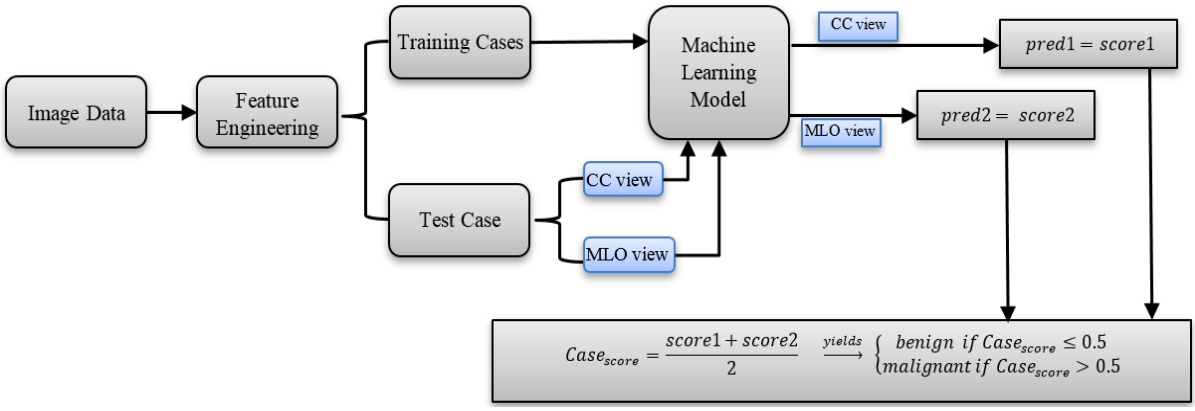


Figure 5-4 Illustration of the overall classification flow of the CAD scheme developed and tested in this study.

Second, we apply the RPA to reduce the dimensionality of image feature space and map to the most efficient feature vector as input features of the SVM model. To demonstrate the potential advantages of using RPA in developing machine learning models, we build and compare 5 SVM models, which using all 181 image features included in the initial feature pool, and embedding 4 other feature dimensionality reduction methods including (1) random

projection algorithm (RPA), (2) principal component analyses (PCA), (3) nonnegative matrix factorization (NMF), and (4) Chi-squared (Chi2).

Third, to increase size and diversity of training cases, as well as reduce the potential bias in case partitions, we use a leave-one-case-out (LOCO) based cross-validation method to train SVM model and evaluate its performance. All feature dimensionality reduction methods discussed in the second step are also embedded in this LOCO iteration process to train the SVM. This can diminish the potential bias in the process of feature dimensionality reduction and machine learning model training as demonstrated in our previous study [121]. When the RPA is embedded in the LOCO based model training process, it helps generate a feature vector independent of the test case. Thus, the test case is unknown to both RPA and SVM model training process. In this way, in each LOCO iteration cycle, the trained SVM model is tested on a truly independent test case by generating an unbiased classification score for the test case. As a result, all SVM-generated classification scores are independent of the training data. In addition, other N-fold cross-validation methods (i.e., $N = 3, 5, 10$) are also tested and compared with LOCO method in the study.

Fourth, since majority of lesions detected in two ROIs from CC and MLO view mammograms, in the LOCO process, two ROIs representing the same lesion will be grouped together to be used for either training or validation to avoid potential bias. After training, ROIs in one remaining case will be used to test the machine learning model that generates a classification score to indicate the likelihood of each testing ROI depicting a malignant lesion. The score ranges from 0 to 1. The higher score indicates a higher risk of being malignant. In addition to the classification score of each ROI, a case-based likelihood score is also generated by fusion of two scores of two ROIs representing the same lesion depicting on CC and MLO view mammograms.

Fifth, a receiver operating characteristic (ROC) method is applied in the data analysis. Area under ROC curve (AUC) is computed from the ROC curve and utilized as an evaluation index to evaluate and compare performance of each SVM model to classify between the malignant and benign lesions. Then, we also apply an operating threshold of T = 0.5 on the SVM-generated classification scores to classify or divide all testing cases into two classes of malignant and benign cases. By comparing to the available ground-truth, a confusion matrix for the classification results is determined for each SVM. From the confusion matrix, we compute classification accuracy, sensitivity, specificity, and odds ratio (OR) of each SVM model based on both lesion region and case. In the region-based performance evaluation, all lesion regions are considered independent, while in the case-based performance evaluation, the average classification score of two matched lesion regions (if the lesions are detected and marked by radiologists in both CC and MLO view) is computed and used. In this study, all pre-processing and feature extraction steps to make the matrix of features are conducted using the MATLAB R2019a package.

5.3 Results

Figure 5-5 shows a malignant case as an example in which radiologists annotate the lesion center in both CC and MLO view mammograms. Based on the marked center, we plot two square areas on two images in which the CAD scheme computes image features. Using the whole feature vector of 181 image features, the SVM-model generates the following classification scores to predict the likelihood of two lesion regions on two-view images being malignant, which are $S_{CCview} = 0.685$, and $S_{MLOview} = 0.291$. The case-based classification score is $S_{Case} = 0.488$. When using the feature vectors generated by the RPA, the SVM-model generates two new classification scores of these two lesion regions, which are $S_{CCview} = 0.817$, and $S_{MLOview} = 0.375$. Thus, the case-based classification score is $S_{Case} = 0.596$. As a result, using the SVM model trained using all 181 image features misclassifies this malignant lesion

into benign when an operating threshold (T = 0.5) is applied, while the SVM model trained using the embedded RPA increases the classification scores for both lesion regions depicting on CC and MLO view images. As a result, it is correctly classified as malignant with the case-based classification score greater than the operating threshold.

Table 5-5 summarizes the performance of using the original features computed in 6 categories to classify between the malignant and benign lesions. As shown in this table, using the group of statistical features yields the highest classification accuracy among 6 categories of features. Figure 5-6 shows a curve indicating the variation trend of the AUC values of the SVM models trained and tested using different number of features (ranging from 50 to 100) generated by the proposed RPA. The trend result indicates that using a reduced feature dimensionality with 80 features, the SVM yields the highest AUC value of 0.84.

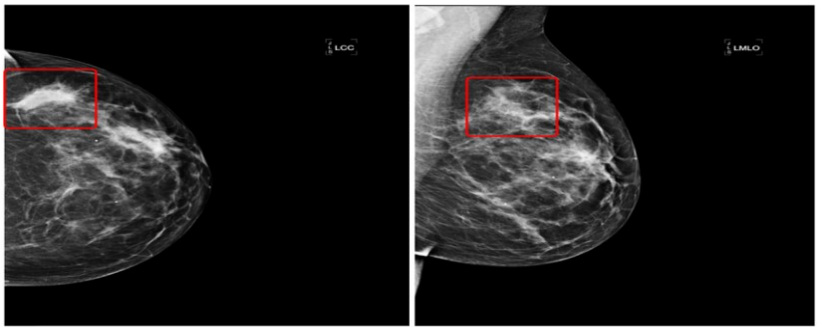


Figure 5-5 A malignant case annotated by radiologists in both CC and MLO views. The annotated mass is squared in each view.

Table 5-5 Accuracy of the SVM models for case-based classification based on six different categories of the original features.

Feature category	Number of features	Accuracy (%)
Statistical features	60	66
GLRLM	7	59
GLDM	24	56
GLCM	44	61
Wavelet-based	26	60
Geometrical based	20	63

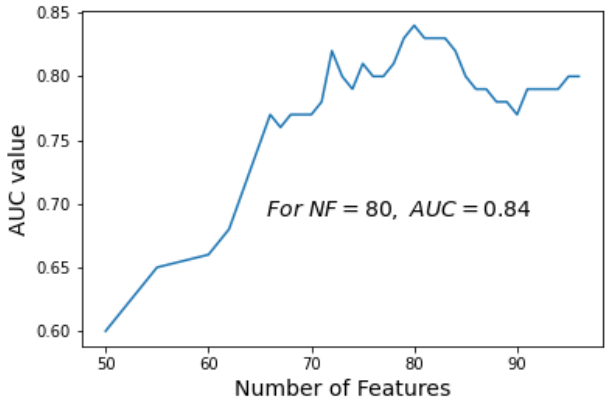


Figure 5-6 A trend of the case-based classification AUC values generated by the SVM models trained using different number of features (NF) generated by the RPA.

Table 5-6 shows and compares the average number of the input features used to train 5 SVM models with and without embedding different feature dimensionality reduction methods, lesion region-based and case-based classification performance of AUC values. When embedding a feature dimensionality reduction algorithm, the size of feature vectors in different LOCO-based SVM model training and validation cycle may vary. Table 5-6 shows that average number of features is reduced from the original 181 features to 100 or less. When using RPA, the average number of features is 80. From both Table 5-6 and Figure 5-7, which show and compare the corresponding AUC values and ROC curves, we observe that an SVM model trained using an embedded RPA feature dimensionality reduction method produces the statistically significantly higher or improved classification performance, including a case-based AUC value of 0.84±0.01 as compared to all other SVM models ($p < 0.05$) including the SVM trained using the initial feature pool of 181 features and other SVM models embedded with other three feature dimensionality reduction methods namely, principal component analyses (PCA), nonnegative matrix factorization (NMF) and Chi-squared (Chi2) in the classification model training process. In addition, the data in Table 5-6 and ROC curves in Figure 5-7 also indicate that the case-based lesion classification yields higher performance than the region-based classification performance, which indicates that using and combining image features computed from two-view mammograms has advantages.

Table 5-6 Summary of the average number of image features used in 5 different SVM models and classification performance (AUC) based on both region and case-based lesion classification. P-value compares the results of each model to the last one (RPA) as the optimal method.

Feature sub-groups	Number of features	AUC	p value
Original features, region based	181	0.72	0.004
Original features, case based	181	0.74	0.005
NMF, region based	100	0.73	0.005
NMF, case based	100	0.77	0.023
Chi2, region based	76	0.73	0.005
Chi2, case based	76	0.75	0.015
PCA, region based	83	0.75	0.011
PCA, case based	83	0.79	0.041
RPA, region based	80	0.78	0.035
RPA, case based	80	0.84	---

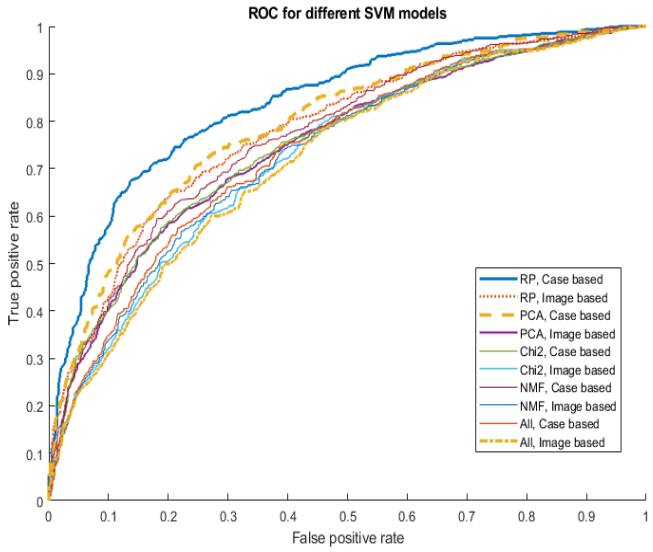


Figure 5-7 Comparison of 10 ROC curves generated using 5 SVM models and 2 scoring (region and case-based) methods to classify between malignant and benign lesion regions or cases.

Table 5-7 presents 5 confusion matrices of lesion case-based classification using 5 SVM-models after applying the operating threshold (T = 0.5). Based on this table, several lesion classification performance indices like sensitivity, specificity, and odds ratio are measured and shown in Table 5-8. This table also shows that the SVM model trained based on the feature vector generated by the RPA yields the highest classification accuracy compared to the other 4

SVM models trained using feature vectors generated based on the other three dimensionality reduction methods or the original feature pool of 181 features.

Table 5-7 Five comparison matrices of case-based lesion classification using five different SVM models to classify benign and malignant cases.

Feature Group	Predicted	Actual Positive	Actual Negative
Original features	Positive	399	212
	Negative	245	631
NMF	Positive	406	173
	Negative	238	670
Chi2	Positive	405	194
	Negative	239	649
PCA	Positive	436	197
	Negative	208	646
RPA	Positive	452	177
	Negative	192	666

Table 5-8 Summary of the lesion case-based classification accuracy, sensitivity, specificity, and odds ratio of using five 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**

Table 5-9 shows and compares the classification results using four different cross-validation methods ($N = 3, 5, 10$ and LOCO). The results show two trends of performance decrease and standard deviation increase (in both AUC and accuracy) as the number of folds decreases from the maximum folds (LOCO) to the smallest folds ($N = 3$). This indicates that using LOCO yields not only the highest performance, but also probably the highest robustness due to the smallest standard deviation.

Table 5-9 Summary of the case-based lesion classification for the proposed method (RPA) under different cross-validation (CV) techniques.

CV	AUC	Accuracy
LOCO	0.84±0.04	75.2±4
10-fold	0.83±0.05	74.0±4
5-fold	0.82±0.07	73.1±5
3-fold	0.80±0.10	70.8±9

Additionally, to assess the reduction of feature redundancy after applying RPA, we create a feature correlation matrix, $corr(i,j)$ with the number of M features. Then, we compute a mean absolute value of the correlation matrix:

$$mean\ of\ correlation = \frac{1}{M \times M} \sum_{i,j=1}^{M} |corr(i,j)| \qquad 5\text{-}12$$

Two mean values of correlation computed from two correlation matrices generated using the feature space (or pools) before and after applying RPA are 0.49 and 0.31, respectively, which indicates that feature correlation coefficients after using RPA are reduced. Thus, using RPA can reduce not only the dimensionality of feature space, but also the redundancy of the feature space.

Last, the computational processing tasks of applying RPA to generate optimal features and train the SVM model are performed using a Dell computer (Processor: Intel(R) Xeon CPU E5-1603 v3, 2.8 GHz, and 16 GB RAM) and Python-based software package. For cross-validation process, we use the Sklearn-model library. For example, in the 10-fold cross-validation, the average computation time to complete one cross-validation iteration is approximately 38.12 seconds.

5.4 Discussion

Mammography is a popular imaging modality used in breast cancer screening and early cancer detection. However, due to the heterogeneity of breast lesions and dense fibro-glandular

tissue, it is difficult for radiologists to accurately predict or determine the likelihood of the detected suspicious lesions being malignant. As a result, mammography screening generates high false-positive recall rates, and the majority of biopsies are approved to be benign [190]. Thus, to help increase breast lesion classification specificity and reduce unnecessary biopsies, developing CAD schemes to assist radiologists more accurately and consistently classifying malignant and benign breast lesions remains an active research topic [191]. In this study, we develop and assess a new CAD scheme of mammograms to predict the likelihood of the detected suspicious breast lesions being malignant. This study has the following unique characteristics as compared to other previous CAD studies reported in the literature.

First, previous CAD schemes of mammograms computed image features from either the segmented lesion regions or the regions with a fixed size (i.e., squared ROIs to cover lesions with varying sizes). Both approaches have advantages and disadvantages. Due to the difficulty to accurately segment subtle lesions with fuzzy boundary, the image features computed from the automatically segmented lesions may not be accurate or reproducible, which reduces the accuracy of the computed image features to represent actual lesion regions. When using the fixed ROIs (including most DL-based CAD schemes [12, 192]), although it can avoid the potential error in lesion segmentation, it may lose and reduce the weight of the image features that are more relevant to the lesions due to the potential heavy influence of irregular fibro-glandular tissue distribution surrounding the lesions with varying sizes. In this study, we tested a new approach that combines image features computed from both a fixed ROI and the segmented lesion region. In addition, comparing to the most of previous CAD studies as surveyed in the previous study, which used several hundreds of malignant and benign lesion regions [8], we assemble a much larger image dataset with 1,847 cases or 2,499 lesion region (including 1,197 malignant lesion regions and 1,302 benign lesion regions). Despite using a much larger image dataset, this new CAD scheme yields a higher classification performance (AUC = 0.84±0.01) as comparing to AUC of 0.78 to 0.82 reported in our previous CAD studies

that using much smaller image dataset (<500 malignant and benign ROIs or images) [12, 78]. Thus, although it may be difficult to directly compare the performance of CAD schemes tested using different image datasets as surveyed in [8], we believe that our new approach to combine image features computed from both a fixed ROI and the segmented lesion region has advantages to partially compensate the potential lesion segmentation error and misrepresentation of the lesions related image features, and enable to achieve an improved or very comparable classification performance.

Second, since identifying a small but adequate and non-redundant image feature vector plays a vital role in CAD development to train machine learning classifiers or models, many feature selection or dimensionality reduction methods have been investigated and applied in previous studies. Although these methods can exclude many redundant and low-performed or irrelevant features in the initial pool of features, the challenge of how to build a small feature vector with orthogonal feature components to represent the complex and non-linear image feature space remains. For the first time, we in this study introduce the RPA to the medical imaging informatics field to develop CAD schemes. RPA is a technique that maximally preserves the distance between the sub-set of points in the lower-dimension space. As explained in the Introduction section, in the lower space under preserving the distance between points, classification is much more robust with a low risk of overfitting. This is not only approved by the simulation or application results reported in previous studies, it is also confirmed by this study. The results in Table 5-6 show that by using the optimal feature vectors generated by RPA, the SVM model yields significantly higher classification performance in comparison with other SVM models trained using either all initial features or other feature vectors generated by other three popular feature selection and dimensionality reduction methods. Using the RPA boosts the AUC value from 0.72 to 0.78 in comparison with the original feature vector in the lesion region-based analysis, and from 0.74 to 0.84 in the lesion case-base evaluation, which also enhances the classification accuracy from 69.3% to 75.2%, and approximately doubling

the odds ratio from 4.85 to 8.86 (Table 5-8). Thus, the study results confirm that RPA is a promising technique applicable to generate optimal feature vectors for training machine learning models used in CAD of medical images.

Third, since the heterogeneity of breast lesions and surround fibro-glandular tissues distributed in 3D volumetric space, the segmented lesion shape and computed image features often vary significantly in two projection images (CC and MLO view), we investigate and evaluate CAD performance based on single lesion regions and the combined lesion cases if two images of CC and MLO views were available and the lesions are detectable on two-view images. Table 5-6 shows and compares lesion region-based and case-based classification performance of 5 SVM models. The result data clearly indicates that instead of just selecting one lesion region for likelihood prediction, it would be much more accurate when the scheme processes and examines two lesion regions depicting on both CC and MLO view images. For example, when using the SVM trained with the feature vectors generated by the RPA, the lesion case-based classification performance increases 7.7% in AUC value from 0.78 to 0.84 as comparing to the region-based performance evaluation.

Last, although the study has tested a new CAD development method using an RPA to generate optimal feature vector and yielded encouraging results to classify between the malignant and benign breast lesions, we realize that the reported study results are made on a laboratory-based retrospective image data analysis process with several limitations. First, although the dataset used in this study is relatively large and diverse, whether this dataset can sufficiently represent the real clinical environment or breast cancer population is unknown or not tested. All FFDM images were acquired using one type of digital mammography machines. Due to the difference of the image characteristics (i.e., contrast-to-noise ratio) between FFDM machines made by different vendors, the CAD scheme developed in this study may not be directly and optimally applicable to mammograms produced by other types of FFDM machines. However, we believe that the concept demonstrated in this study is valid. Thus, the similar CAD

schemes can be easily retrained or fine-tuned using a new set of digital mammograms acquired using other different types of FFDM machines of interest. Second, in this retrospective study, the image dataset has a higher ratio between the malignant and benign lesions, which is different from the false-positive recall rates in the clinical practices. Thus, the reported AUC values may also be different from the real clinical practice, which needs to be further tested in future prospective clinical studies. Third, in the initial pool of features, we only extracted a limited number of 181 statistics, textural and geometrical features, which are much less than the number of features computed based on recently developed radiomics concept and technology [171, 172]. Thus, more texture features can be explored in future studies to increase the diversity of the initial feature pool, which may also increase the chance of selecting or generating more optimal features. Additionally, many deep transfer learning models have been recently tested as feature extractors in the medical imaging field, which produce a much larger number of features than the radiomics approaches. Thus, whether using RPA can also help significantly reduce the dimensionality of these feature extractors to more effectively and robustly train or build the final classification layer of the DL models should be investigated in future studies.

6 Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms

6.1 Introduction

From the end of 2019, a new coronavirus namely COVID-19, was confirmed in human bodies as a new category of diseases that cause dangerous respiratory problems, heart infection, and even death. To more effectively control COVID-19 spread and treat patients to reduce mortality rate, medical images can play an important role [193, 194]. In current clinical practice, chest X-ray radiography and computed tomography (CT) are two imaging modalities to detect COVID-19, assess its severity, and monitor its prognosis (or response to treatment). Although CT can achieve higher detection sensitivity, chest X-ray radiography is more commonly used in clinical practice due to the advantages, including low cost, low radiation dose, easy-to-operate and wide accessibility in the general or community hospitals [195, 196]. However, pneumonia can be caused by many different types of viruses and bacterial. Thus, it may be time-consuming and challenging for general radiologists in the community hospitals to read a high volume of chest X-ray images to detect subtle COVID-19 infected pneumonia and distinguish it from other community-acquired non-COVID-19 infected pneumonia. It is because there are many similarities between pneumonia infected by COVID-19 and other types of viruses or bacteria. Thus, this is a clinical challenge faced by the radiologists in this pandemic [197].

To address this challenge, developing computer-aided detection or diagnosis (CAD) schemes based on medical image processing and machine learning has been attracting broad research interest, which aims to automatically analyze disease characteristics and provide

radiologists valuable decision-making supporting tools for more accurate or efficient detection and diagnosis of COVID-19 infected pneumonia. To this aim, studies may involve the following steps of preprocessing images, segmenting regions of interest (ROIs) related to the targeted diseases, computing and identifying effective image features, and building multiple-feature fusion-based machine learning models to detect and classify cases. For example, one study [198] computed 961 image features from the segmented ROIs depicting chest X-ray images. After applying a feature selection algorithm, a KNN classification model was built and yielded an accuracy of 96.1% to classify between COVID-19 and non-COVID-19 cases.

However, due to the difficulty in identifying and segmenting subtle pneumonia-related disease patterns or ROIs on chest X-ray images, recent studies have demonstrated that developing CAD schemes based on deep learning algorithms without segmentation of suspicious ROIs and computing handcrafted image features is more efficient and reliable than the use of the classical machine learning methods. As a result, many deep learning models have been reported recently in the literature to detect and classify COVID-19 cases [195, 199-207]. Although some deep learning convolution neural network (CNN) models are applied to CT images [199, 200], more studies applied CNN models to detect and classify COVID-19 cases using chest X-ray images. They include different existing CNN models (i.e., Resnet50 [195, 201], MobileNetV2 [202], CoroNet [203], Xception+ResNet50V2 [204]) and several new special CNN models (i.e., DarkCovidNet [205], COVID-Net [206] and COVIDX-Net [207]). These studies used different image datasets with a varying number of COVID-19 cases (i.e., from 25 to 224) among the total number of cases from 50 to 11,302. The reported sensitivity to detect COVID-19 cases ranged from 79.0% to 98.6%.

Despite the promising results reported in previous studies, many issues have not been well investigated regarding how to train deep learning models optimally. For instance, whether applying image preprocessing algorithms can help to improve the performance and robustness of the deep learning models. To better address some of the challenges or technical issues, we in

this study develop and test a new DL-based CAD scheme of chest X-ray radiography images. The scheme can detect and classify images into 3 classes namely, COVID-19 infected pneumonia, the other community-acquired non-COVID-19 infected pneumonia, and normal (non-pneumonia) cases. The hypothesis in this study is that instead of directly using the original chest X-ray images to train deep learning models, we can apply image processing algorithms to remove the majority of diaphragm regions, normalize image contrast and reduce image noise, and generate a pseudo color image to feed in 3 input channels of the existing deep learning models that were pre-trained using color (RGB) images in the transfer learning process. It may help significantly improve model performance and robustness in detecting COVID-19 cases and distinguishing them from other community-acquired non-COVID-19 infected pneumonia cases. To test this study hypothesis and demonstrate the potential advantages of new approaches, we assemble a relatively large chest X-ray image dataset with 3 class cases. Then, we select a well-trained VGG16 based CNN model as a transfer learning model used in our CAD scheme. The details of the study design and data analysis results are reported in the following sections of this article.

6.2 Materials and Methods

6.2.1 Dataset

In this study, we utilize and assemble a dataset of chest X-ray radiography (CXR) images that are acquired from several different publicly available medical repositories [208-212]. These repositories were initially created and examined by the Allen Institute for AI in partnership with the Chan Zuckerberg Initiative, Georgetown University's Center for Security and Emerging Technology, Microsoft Research, and the National Library of Medicine - National Institutes of Health, in coordination with The White House Office of Science and Technology Policy. Specifically, the dataset used in this study includes 8,474 2D X-ray images in the posteroanterior (PA) chest view. Among them, 415 images depict with the confirmed COVID-

19 disease, 5,179 with other community-acquired non-COVID-19 infected pneumonia, and 2,880 normal (non-pneumonia) cases.

6.2.2 Image Preprocessing

Figure 6-1 shows examples of three chest X-ray images acquired in three classes of normal, community-acquired non-COVID-19 infected pneumonia and COVID-19 pneumonia cases (from top to bottom). It shows that the bottom part of the images includes a diaphragm region with high-intensity (or bright pixels), which may have a negative effect on distinguishing and quantifying lung disease patterns using deep learning models. Hence, an image pre-processing algorithm is applied to identify and remove diaphragm regions. Specifically, the algorithm detects the maximum (the brightest - V_{max}) and minimum (the darkest - V_{min}) pixel value of the image, then uses a threshold $T = V_{min} + 0.9 \times (V_{max} - V_{min})$ to segment the original image into a binary image, as shown in Figure 6-1(b). Next, after labeling all connected regions in the binary image, the CAD scheme detects the biggest region, fills the holes in this region, and deletes all other small regions (if any), as shown in Figure 6-1(c). This detected region locates in the diaphragm. Then, morphological filters are applied to smooth the boundary of the region, as shown in Figure 6-1(d). Last, the processed binary image is mapped back to the original image, the CAD scheme removes overlapped pixels in the corresponding locations on the original image, as shown in Figure 6-1(e). Images after this step are named (I_p).

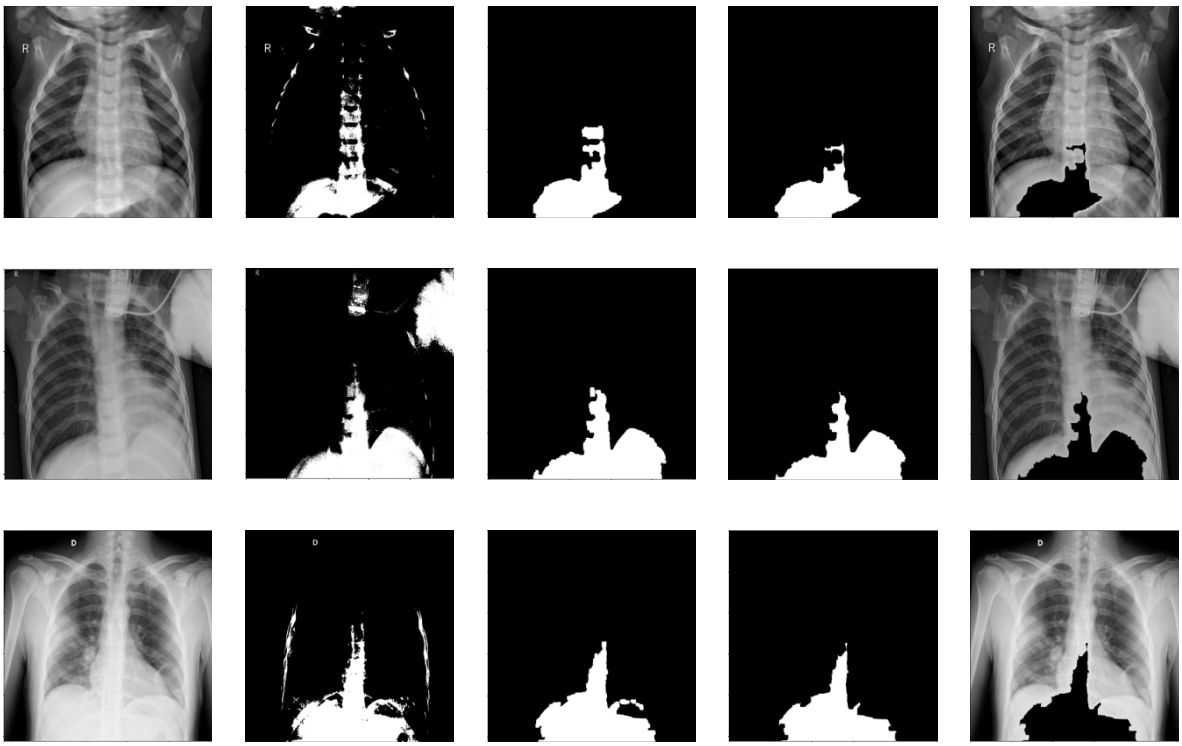


Figure 6-1 Example of chest X-ray images in three classes (top – normal, middle – community-acquired non-COVID-19 pneumonia, and bottom – COVID-19 infected pneumonia case). The figure also shows (a) the original Images, (b) the binary images after threshold, (c) images after selecting the biggest segmented region, (d) images after applying morphological filtering and (e) the original image after removing the majority part of diaphragm region (I_p).

In the next step, we convert the segmented grayscale images (I_p) to 3-channel images suitable for fine-tuning an existing CNN model pre-trained using color (RGB) images. To do so, we apply an image noise filtering method and a contrast normalization method to preprocess the image after removing the diaphragm region. First, since the X-ray images often include additive noise, we apply a bilateral low-pass filter (BF) to I_p. This filter is a non-linear filter and highly effective at noise removal while preserving textural information compared to the other low pass filters. In other words, this filter analyzes intensity values locally and considers the intensity variation of the local area to replace the intensity value of each pixel with the averaged intensity value of the pixels in the local area. To calculate the weights, we apply a Gaussian low-pass filter in the space domain. This step generates a noise-reduction image.

Based on our experimental results, we select the following parameters in the bilateral filtering ($d = 9$ and $\sigma = 75$). Second, chest X-ray images may have different image contrast or brightness due to the difference in patient body size and/or variation of X-ray dose. To compensate such a potentially negative impact, we apply a histogram equalization (HE) method to normalize I_p images. This filter can enhance lung tissue patterns and characteristics associated with COVID-19 infection. Then, as shown in Figure 6-2, three preprocessed images namely, I_p, $I_b = BF(I_p)$ and $I_{eq} = HE(I_p)$ form a pseudo color image that is fed into 3 input (RGB) channels of the CNN model.

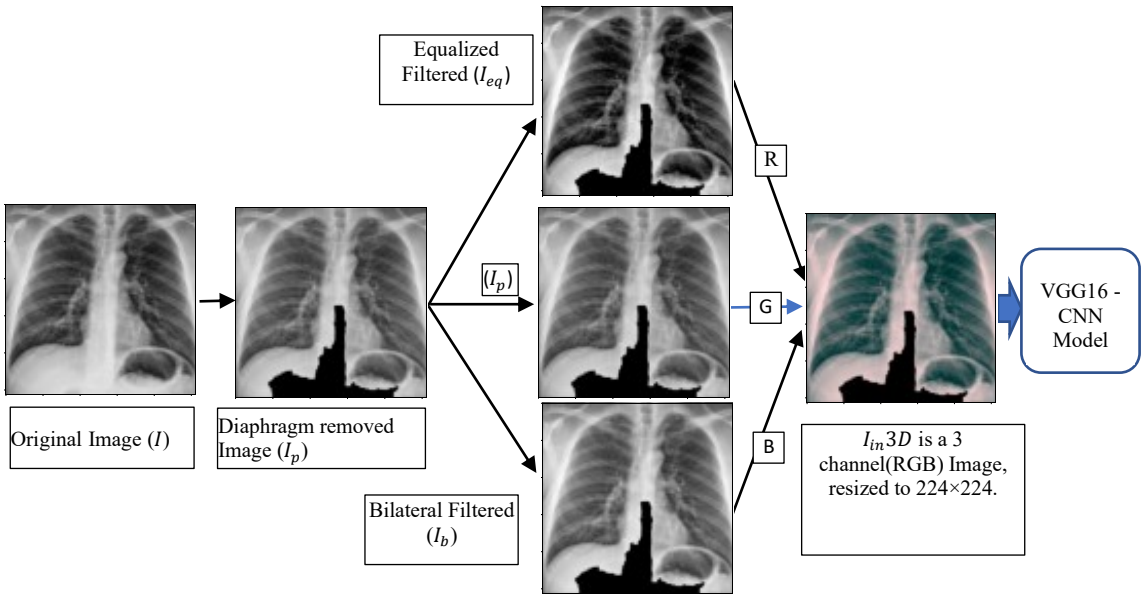


Figure 6-2 A flow diagram to illustrate image pre-processing steps to generate input of a CNN model, where (I) is the original Image in the dataset. (I_p) is the diaphragm removed image. (I_eq) is an image after applying histogram equalization on (I_p), and (I_b) is an image after applying bilateral filtering on (I_p). Three images (I_p), (I_b), and (I_eq) are fed into three channels of the CNN model to simulate the RGB image.

6.2.3 Transfer Learning

In this study, we adopt a transfer learning approach since the previous studies have shown in order to avoid either overfitting or underfitting consequences using a small training dataset, a better approach is to take advantage of a CNN initially trained using a large-scale dataset

[213]. Currently, many CNN models have been previously developed and are available for different engineering applications. In this study, we select a VGG16 model, which was pre-trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) using a large dataset with 14 million images [214]. VGG16 model won the first place on the image localization task and second place on the image classification task in the 2014 ILSVRC challenge [215]. As shown in Figure 6-3, the VGG16 model has 13 convolutions, 5 max-pooling, and 3 fully connection layers in 6 blocks, which include over 138 million trainable parameters.

In our transfer learning, the weights between all connected nodes in front or low layers of the VGG16 based CNN model maintain unchanged (blocks 1 to 5 as shown in Figure 6-3). Next, block 6 in the model is modified by replacing with one flatten layer and two fully connected layers, which include 256 and 128 nodes, respectively. In these layers, the rectified linear unit (ReLU) [216] is used as their activation function. Then, all trainable weights in all connection nodes of the whole modified VGG16 model are fine-tuned using chest X-ray image data. In this fine-tuning process, a small learning rate (learning rate = 10^{-5}) is used to make a small variation to the pre-trained parameters. In this way, we will preserve the valuable parameters as much as possible by avoiding dramatic changes on the pre-trained parameters and let the model learn the special characteristics of chest X-ray images. Finally, in the last classification layer, Softmax is used as the activation function. As a result, a new transfer learning model is built to fulfill a three-class classification task. The complete CNN model is compiled with Adam [217] optimizer with a batch size of 4, max epoch = 200, initial learning rate = 10^{-5}, and monitoring validation loss for reducing the learning rate every 5 epochs with a factor of 0.8. Table 6-1 shows the complete architecture of the transfer learning VGG16 model built in this study.

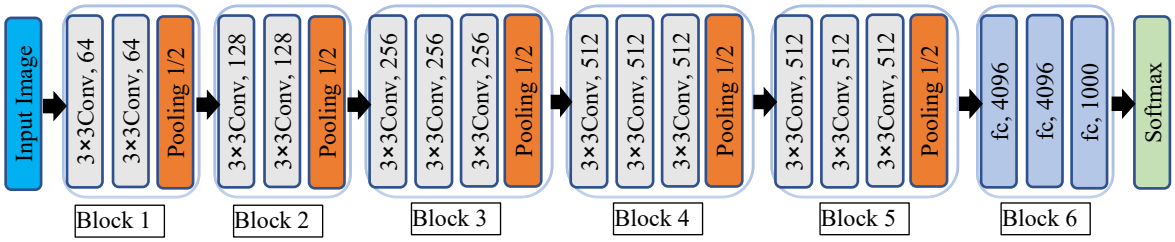


Figure 6-3 Illustration of the architecture of VGG16 based CNN model.

Table 6-1 The architecture of the new VGG16 model after transfer learning with new layers (19 to 22).

Number	Layer	Size	Activation
0	Input Image	$224 \times 224 \times 3$	---
2	2× Convolution (3×3)	$224 \times 224 \times 64$	ReLu
3	Max Pooling	$112 \times 112 \times 64$	ReLu
5	2× Convolution (3×3)	$112 \times 112 \times 128$	ReLu
6	Max Pooling	$56 \times 56 \times 128$	ReLu
9	3× Convolution (3×3)	$56 \times 56 \times 256$	ReLu
10	Max Pooling	$28 \times 28 \times 256$	ReLu
13	3× Convolution (3×3)	$28 \times 28 \times 512$	ReLu
14	Max Pooling	$14 \times 14 \times 512$	ReLu
16	3× Convolution (3×3)	$14 \times 14 \times 512$	ReLu
18	Max Pooling	$7 \times 7 \times 512$	ReLu
19	Flattening	25088	---
20	Fully Connected	256	ReLu
21	Fully Connected	128	ReLu
22	Fully Connected	3	SoftMax

6.2.4 Model Training and Testing

First, the original chest X-ray image has 1,024×1,024 pixels, while the VGG16 model was pre-trained using images of 224×224 pixels. Thus, each chest X-ray image is down-sampled to 224×224 pixels to fit the VGG16 model. Then, for training and evaluating the proposed VGG16 based transfer learning CNN model, we randomly split the entire image dataset of 8,474 cases

into 3 independent subsets of training, validation, and testing. Overall, 10% of cases (848) are assigned to the test subset. On the remaining 7,626 cases, 10% cases are assigned to the validation subset (757), while 90% cases (6,869) are formed as the training subset. To maintain the same case partition ratios for three classes of COVID-19 infected pneumonia, other community-acquired non-COVID-19 infected pneumonia and normal cases, the case partition or assignment is done on three classes independently. Table 6-2 shows the number of cases in each subset.

Table 6-2 Distribution of cases in three subsets.

Image Data Subset	Training	Validation	Testing
COVID-19 cases	366	37	42
Other pneumonia cases	4,201	460	518
Normal cases	2,332	260	288
Total number of cases	6,899	757	848

Second, there are different available techniques to deal with imbalanced data [218][24]. In this study, the class weight technique, as one possible way, is applied during training to reduce the potential consequences of imbalanced data. In the class weigh technique, we adjust weights inversely proportional to class frequencies in the input data [219]. The weight, ω_i in class i is computed using the following equation.

$$\omega_i = {(Total\ number\ of\ cases)}\Big/{(number\ of\ classes)\ \times (number\ of\ cases\ in\ class(i))} \quad 6\text{-}1$$

The weights of the classes are utilized while fitting the model. Hence, in the loss function, we assign higher values to the instances of smaller classes. Therefore, the calculated loss will be a weighted average, where the weights of each sample corresponding to each class during loss calculation are specified with ω_i.

Additionally, in the training data of minority cases (COVID-19 cases), a common augmentation technique [220] is applied to increase the training sample size. First, using shearing factors (≤0.2), image intensity is sheared based on the shearing angle in a counter-clockwise direction. Second, using zooming factors (≤0.2), images are randomly magnified. Third, using rotation factors (within $\pm 20^{\circ}$), images are randomly rotated. Fourth, using a shift factor (≤0.2), images are randomly shifted in 4 directions (up, left, down right). Last, images are flipped horizontally in a random base to generate as much augmented data as possible.

Multiple iteration or epochs are applied to train the VGG16 based CNN model. The model is first trained using the data in the training subset and validated using the validation subset. During the training process, the optimizer tries to force the architecture to learn more and more information to reduce the performance gap between training and validation. To control overfitting and maintain training efficiency, we limit model training epochs to 200. Hence, at the end of 200 training epoch, the trained model is saved and then tested using the data in the testing subset, which does not involve in the model training and validation process.

To reduce the risk of potential bias in data partition into three subsets of training, validation, and testing, we repeat this model training and testing process three times by randomly dividing all cases into training, validation, and testing subsets three times using the same case ratios or numbers as shown in Table 6-2. In addition, during these three times of case partition, the cases assigned to the validation and testing subsets are totally different (no duplication). Three trained models are tested using totally different testing cases. Thus, the total number of testing cases increases (as shown in Table 6-2) to 2,544 (848× 3). Figure 6-4 shows a schematic diagram that illustrates the complete architecture of this VGG16 transfer learning CNN model, as well as the training, validation, and testing phase.

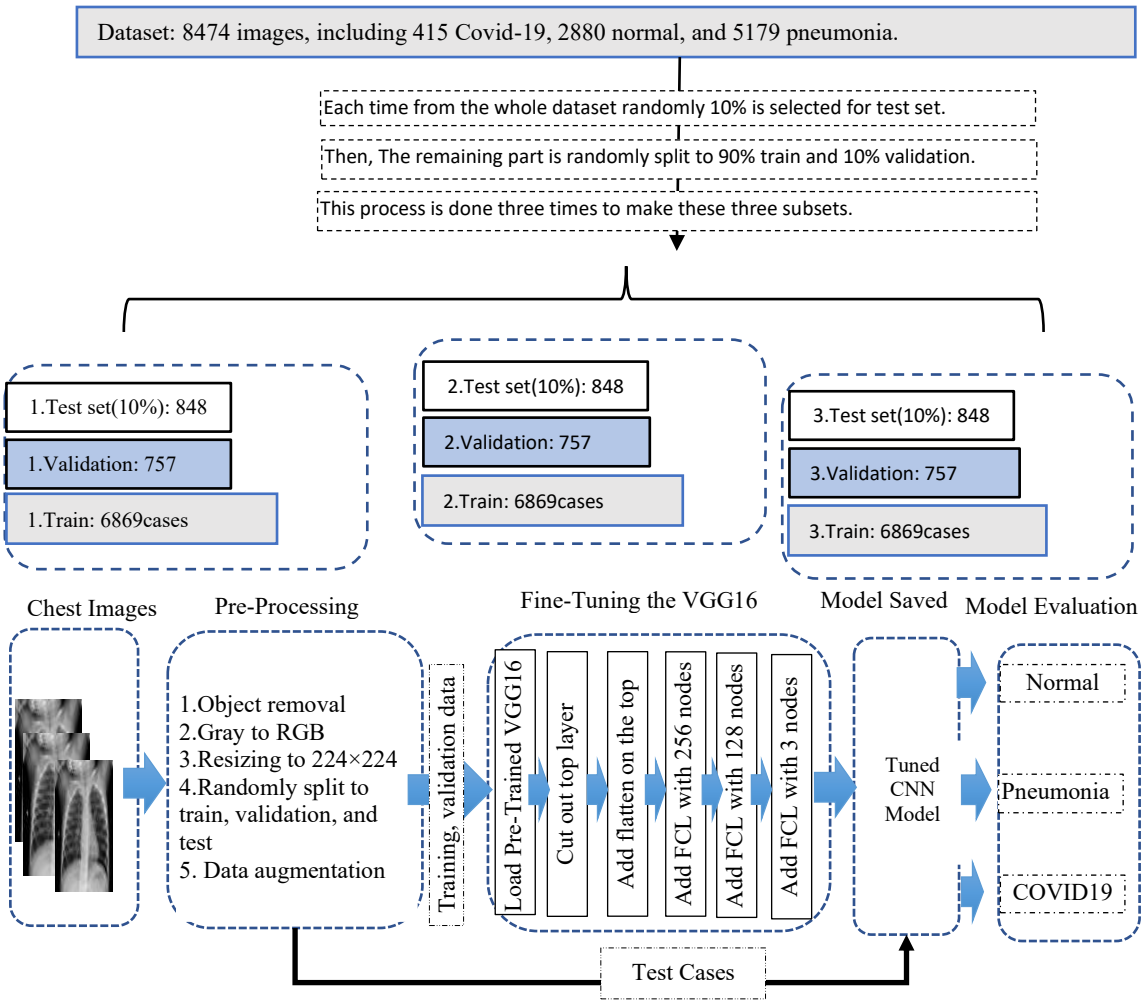


Figure 6-4 schematic representing training and validation phase of the proposed scheme.

6.2.5 Performance Assessment

We perform experiments to analyze two different accuracies. The first one is accuracy for a three-class classification to distinguish between COVID-19 infected pneumonia, community-acquired pneumonia, and normal (non-pneumonia) cases. We compute accuracy values in detecting images in 3 classes. We also calculate (1) a macro averaging, which is the average of 3 accuracy values of 3 classes without considering the proportion of the number of the cases in each class ($A_{mac} = (A_1 + A_2 + A_3)/3$), and (2) a weighted averaging, which is the weighted average of 3 accuracy values weighted with respect to the proportion of the classes ($A_w =$

$w_1A_1 + w_2A_2 + w_3A_3$), where A_1, A_2, A_3 are accuracy values of 3 classes, while w_1, w_2, w_3 are weighting factors of 3 classes representing the ratios of cases in 3 classes. Then, for the three-class classification, a confusion matrix is generated from which several evaluation indices, including precision, recall, F1-score, and Cohen's Kappa [221] values are computed to evaluate CAD performance. The value of Cohen's kappa coefficients (ranging from zero and one) indicates the possibility of the predicted results occurring by chance. The lower Kappa value shows the more randomness of the results, while the higher value shows a better similarity and higher robustness.

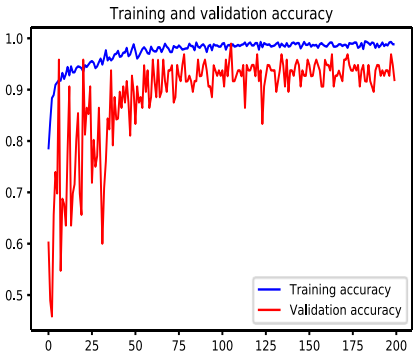
The second accuracy evaluation refers to the classification between the COVID-19 and non-COVID19 cases (including both normal and community-acquired pneumonia cases). In this circumstance, we compute true positive (TP) for the cases correctly identified as COVID-19, false negative (FN) for the COVID-19 cases being incorrectly classified as normal or community-acquired pneumonia cases, true negative (TN) for the cases correctly identified as non-COVID-19 cases, and false positive (FP) for the normal and community-acquired pneumonia cases being incorrectly classified as COVID-19 by the CNN model. Then, the accuracy, sensitivity, specificity, recall, and F1-scores of model classification are computed and tabulated.

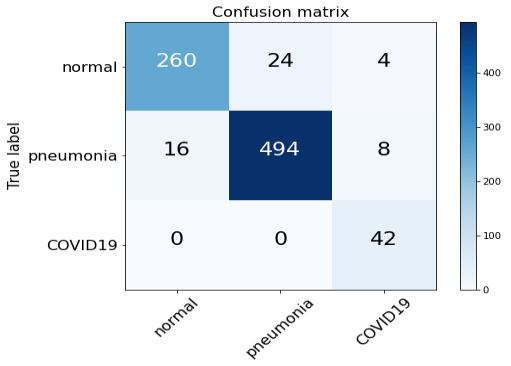
6.3 Results

Figure 6-5(a-c) presents trend curves of training and validation accuracy of the new transfer learning VGG16 based CNN model in three experiments using different training and validation subsets in the left column. Then, by applying the trained models on the corresponding testing subsets, three confusion matrices of the models on the testing subsets are shown in the right column. All three curves show that as the increase of training iteration epochs during the training process, the prediction accuracy of the validation subset varies greatly (with big oscillation) initially, and then gradually converges to a higher accuracy level with much small

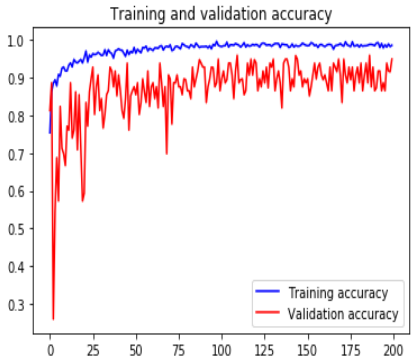
oscillation. Thus, for all three subsets after epoch 75, validation accuracy is following the training accuracy, which indicates that learning is happening during different epochs. The trend graph also shows that the proposed technique does not suffer significant overfitting or underfitting in our transfer learning model. Then, by combining three confusion matrixes of the three independent testing subsets, as shown in the second column of (a-c), Figure 6-5(d) displays a combined 3-class confusion matrix of 2,544 (3×848) cases.

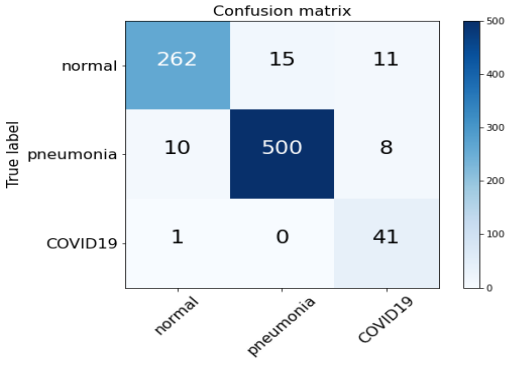
a)



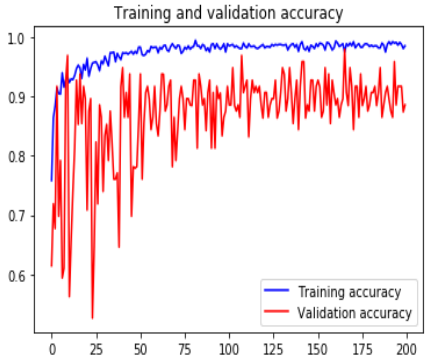


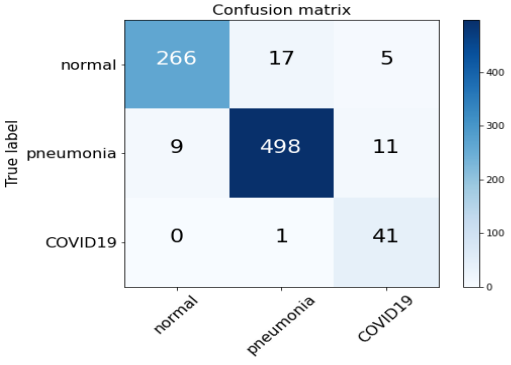
b)





c)





d)

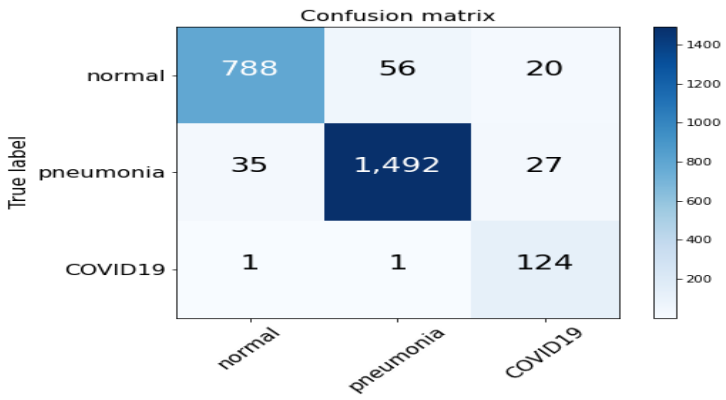


Figure 6-5 (a-c) Left column show three sets of performance curves applying to the training and validation subsets for 3 experiments in 200 training epochs, respectively. The horizontal axis shows the number of epochs, and the vertical axis shows the accuracy. The right column show three confusion matrices of the corresponding testing results are shown on the right. (d) A combined confusion matrix of applying three trained models to three independent testing data subsets with total 2,544 cases.

First, based on three confusion matrices as shown in Figure 6-5(a-c), the overall 3-class classification accuracy levels are 93.9% (796/848), 94.7% (803/848), and 94.9% (805/848), respectively. The difference is approximately 1%. Then, based on the confusion matrix of the combined data as shown in Figure 6-5(d), we compute the precision, recall rate, F1-score, and prediction accuracy of the new transfer learning VGG16 based CNN model, as shown in Table 6-3. Among 2,544 testing cases, 2,404 are correctly detected and classified into 3 classes. The overall accuracy is 94.5% (2,404/2,544) with 95% confidence interval of [0.93 ,0.96]. In addition, the computed Cohen's kappa coefficient is 0.89, which confirms the reliability of the proposed approach to train this new deep transfer learning model to do this classification task.

Table 6-3 Classification report of the proposed method.

	Precision	Recall	F1-score	Support cases
Normal	0.96	0.91	0.93	864
Other Pneumonia	0.96	0.96	0.96	1,554
COVID19	0.73	0.98	0.84	126
Accuracy	---	---	0.95	2,544
Macro avg	0.88	0.95	0.91	2,544
Weighted avg	0.95	0.94	0.94	2,544

To further evaluate the performance of our CAD scheme in detecting the COVID19 infected pneumonia cases using chest X-ray images, we place both normal and community-acquired pneumonia images into the negative class and COVID-19 infected pneumonia cases into the positive class. Combining the data in the confusion matrix, as shown in Figure 6-5(d), the CAD scheme yields 98.4% detection sensitivity (124/126) and 98.0% specificity (2,371/2,418). The overall accuracy is 98.1% (2,495/2,544).

Next, Table 6-4 shows and compares (1) confusion matrixes generated by four models trained and tested using different input images and three data subsets generated from the data partition, as well as (2) overall classification accuracy and 95% confidence intervals. The results indicate that (1) without using the data augmentation technique, the model accuracy on data of the testing subset drops to 82.3% with the kappa score of 0.71. (2) Without applying image preprocessing and directly feeding the original chest X-ray images into the VGG16 based CNN model ("simple model"), classification accuracy is 88.0% with a Cohen's kappa score of 0.75. (3) Using image filtering and pseudo color images without removing the majority part of diaphragm regions, the "filter-based model" yields 91.2% accuracy and a Cohen's kappa score of 0.83. All three models yield lower classification accuracy than the proposed model involving data augmentation technique and two steps of image preprocessing.

In addition, Table 6-5 compares our transfer learning VGG16 based CNN model and 10 state-of-art models recently reported in the literature to detect and classify COVID-19 cases. The table shows the number of cases in the training and testing data subsets, imaging modality (CT or X-ray radiography), and reported classification performance, including either 3-class or 2-class classification for these studies. Although the reported performance of these studies cannot be directly compared due to the use of different image dataset and testing methods, the presented data clearly demonstrate that our model is tested using a relatively large dataset and yields very comparable classification performance as comparing to the state-of-art models developed and tested in this research field.

Table 6-4 Confusion matrix of four CNN models on X-ray Images. 95% confidence interval (CI) for the accuracy is shown in the last column.

			Normal	**Pneumonia**	**COVID19**	**Accuracy**	**95% CI**
Proposed Model	True Label	Normal	788	56	20	94.5%	[0.93,0.96]
		Pneumonia	35	1492	27		
		COVID 19	1	1	124		
Filter-based model		Normal	750	89	25	91.2%	[0.90,0.92]
		Pneumonia	64	1452	38		
		COVID19	2	6	118		
Simple model		Normal	701	123	40	88.0%	[0.86,0.89]
		Pneumonia	72	1431	51		
		COVID19	6	13	107		
No-augmentation		Normal	653	158	53	82.3%	[0.80,0.84]
		Pneumonia	124	1346	74		
		COVID19	8	23	95		

Table 6-5 Comparison accuracy results of the proposed method with the other deep learning methods on COVID-19 diagnosis.

Approach	Data Type	Cases number (including COVID-19 cases)	Method utilized	2 classes accuracy (%)	3 classes accuracy (%)	COVID-19 detection Sensitivity (%)
Narin et al. **[195]**	X-ray	100 (50)	ResNet50	98.0	---	96.0
Sethy et al. **[201]**	X-ray	50 (25)	ResNet50+SVM	95.4	---	97.0
Ioannis et al. **[202]**	X-ray	1,427 (224)	MobileNetV2	96.7	93.5	98.6
Wang et al. **[199]**	CT	237 (119)	M-Inception	82.9	---	81.0
Tulin et al. **[205]**	X-ray	1,127 (127)	DarkCovidNet	98.08	87.02	90.6
Khan et al. **[203]**	X-ray	221 (29)	CoroNet (Xception)	98.8	94.52	95.0
Rahimzadeh & attar **[204]**	X-ray	11,302 (31)	Xception+ ResNet50V2	99.5	91.4	80.53

Wang et al. **[206]**	X-ray	300 (100)	COVID-Net	96.6	93.3	91.0
Ying et al. **[200]**	CT	57 (30)	DRE-Net (ResNet50)	86	---	79.0
Hemdan et al. **[207]**	X-ray	50 (25)	COVIDX-Net	90	---	----
Our new method	X-ray	2,544 (126)	VGG16	98.1	94.5	98.4

6.4 Discussion

In this study, we developed and tested a novel deep transfer learning CNN model to detect and classify chest X-ray images depicting COVID19 infected pneumonia. This study has several unique characteristics as compared to the previously reported studies in this field and produces several new interesting observations. First, since the deep learning CNN model includes a considerable number of parameters that need to be trained and determined, a large and diverse image dataset is required to produce robust results [95]. Although we used a relatively large image dataset of 8,474 chest X-ray images, the dataset is unbalanced in 3 classes of images, and the number of the COVID-19 infected pneumonia cases (415) remains small. Thus, in order to build a robust deep learning model, we apply a class weight technique during the training process and select a well-trained VGG16 model and apply a transfer learning approach. Specifically, the original VGG16 model includes over 138 million parameters. These parameters have been trained and determined using a large ImageNet database of over 14 million images. It is difficult to train so many parameters from scratch robustly using a dataset of 8,474 images. Thus, we retrain or fine-tune the pre-trained VGG16 (as shown in Figure 6-3) to reduce the overfitting risk. Study results demonstrate that this transfer learning approach can yield higher performance with the overall accuracy of 94.5% (2,404/2,544) in the classification of 3 classes and 98.1% (2,495/2,544) in classifying cases with and without COVID-19 infection, as well as the high robustness with a Cohen's kappa score of 0.89.

Second, unlike the regular color photographs, chest X-ray images are gray-level images. Thus, in order to fully use the pre-trained VGG16-based CNN model, we generate two new

gray-level images. Then, instead of applying the original chest X-ray image to the CNN model directly, 3 different gray-level images are fed into 3 input (RGB color) channels of the CNN model. Specifically, we apply a bilateral low-pass filter to generate a noise-reduced image and a histogram equalization method to generate a contrast normalized image. Comparing two approaches of using only original chest X-ray image and 3 different images to generate a pseudo color image as an input image to the CNN model, our study results show that using a pseudo color image approach, overall classification accuracy increases 3.6% from 91.2% to 94.5%, and Cohen's kappa score increases 7.2% from 0.83 to 0.89, respectively. The results demonstrate the advantage of using our new approach to fully use 3 input channels of the CNN model pre-trained using color images because these two filtered gray-level images contain additional information, which can enhance image classification capability.

Third, since in the area of medical imaging, generally, disease's patterns are not comparable to the other existing patterns in the image, preprocessing steps are noteworthy [83]. Hence, we apply an image preprocessing algorithm to automatically detect and remove the majority part of the diaphragm region from the chest X-ray images. Comparing the approaches with and without removing the diaphragm regions, the classification performance of the CNN model changes from 94.5% to 88.0% and 0.89 to 0.75 for the overall classification accuracy and Cohen's kappa coefficients, respectively, which indicates a 7.4% increase in the overall classification accuracy and 18.7% increase in Cohen's kappa coefficient by removing the majority of diaphragm regions. Thus, although skipping segmentation of the suspicious disease regions of interest is one important characteristic of deep learning, our study demonstrates that applying an image preprocessing and segmentation algorithm to remove irrelevant regions on the image can also play an important role in increasing the performance and robustness of deep learning models.

In addition, we observe and confirm that applying data augmentation in the training data is also essential. Without data augmentation to increase training dataset size, the overall

classification accuracy of the CNN model significantly reduces to around 82.3%. In summary, we in this paper present a new deep transfer learning model to detect and classify the COVID-19 infected pneumonia cases, as well as several unique image preprocessing approaches to optimally train the deep learning model using the limited and unbalanced medical image dataset. The similar learning concept and image preprocessing approaches can also be adopted to develop new deep learning models for other medical images to detect and classify other types of diseases (i.e., cancers [222, 223]).

Despite encouraging results, this study also has limitations. First, although we used a publicly available dataset of 8,474 cases, including 415 COVID-19 cases, due to the diversity or heterogeneity of COVID-19 cases, the performance and robustness of this CAD scheme need to be further tested and validated using other large and diverse image databases. Second, this study only investigates and tests two image preprocessing methods to generate two filtered images, which may not be the best or optimal method. New methods should also be investigated and compared in future studies. Third, to further improve model performance and robustness, it also needs to develop new image processing and segmentation algorithms to more accurately remove the diaphragm and other regions outside lung areas in the images. Therefore, more research work is needed to overcome these limitations in the future studies.

7 Conclusion and future work

7.1 Summary

The purpose of developing CAD schemes is to assist physicians (i.e., radiologists) to more accurately interpret medical imaging findings and reduce inter-reader variability [224]. In developing CAD schemes, ML plays an essential role because it is widely used to identify effective image features from complex datasets and optimally integrate them with the classifiers, which aims to assist the clinicians to more accurately detect early disease, classify disease types and predict disease treatment outcome. To present a robust and reliable ML-based CAD scheme, optimizing the performance of three main stages is worthy of consideration. These stages are (1) segmentation for a proper feature extraction domain, and extract appropriate features, (2) feature optimization and data reduction techniques, and (3) optimizing parameters of machine learning/deep learning models to improve CAD performance.

Although it is possible to compute many features from an image, having too many features can lead to overfitting rather than learning the actual basis of a pattern. Besides, in the process of the feature extraction step, some data-related issues refer to the quality of the data, and the preprocessing steps are needed to make them more suitable for ML. Data quality is affected by the issues including the presence of noise, outliers, missing or duplicate data, and data that is biased-unrepresentative. Several different techniques and strategies exist in the medical imaging informatics field, relevant to data preprocessing that focuses on modifying the data for better fitting in a specific ML method [225].

Besides initial preprocessing utilized on the medical images to make them more appropriate for feature extraction by methods like image quality enhancement, and ROI segmentation, it is possible to post-process the extracted features and optimize them for the learning process. The process of selecting a subset of features appropriate to make the best

predictions is known as feature selection [17]. There are many advantages regarding dimensionality reduction and feature selection when the datasets have a large number of features. ML algorithms work better and more robust when the dimensionality of feature space is lower. Additionally, the reduction of dimensionality can exclude irrelevant features, diminish noise and redundant features, and generate more robust learning models due to fewer features. In general, if the dimensionality reduction comes by selecting new features which are a subset of the old ones, it is known as feature selection. On the other side, if the dimensionality reduction comes with combining the initial set of features to regenerate a new sub-set of features is known as a dimensional reduction for the optimized set of features [18].

In addition to selecting optimal features, optimizing classifier parameters is also essential to build a robust classifier. In [23], researchers have presented the study to show that even making a change in a classifier's settings after optimizing the feature domain can change the accuracy of the scheme. Hence, some techniques that can search in classifiers' parameter domain to find the optimal groups of them may present CAD schemes with enhanced performance. This parameter optimization process can either utilize on classical machine learning systems or deep learning systems as a new generation of AI to acquire optimal values for their hyperparameters [226].

In the previous chapters of my dissertation, I developed and presented several CAD schemes as solutions to some of these challenges that were evaluated for different purposes, from the prediction of cancer risk scores to malignancy detection and COVID-19 classification.

In chapter 3, the idea is mainly around the first critical stage of any robust ML-based CAD scheme. I explored a new approach to develop a unique case-based CADx scheme based on the detection, computation, and analysis of globally asymmetrical image features computed from two bilateral images of left and right breasts and assessed its performance using a relatively large image dataset of 1,959 cases. Thus, this new CADx scheme is a multiply image-based

scheme that integrates image feature differences computed from 4 view images, making it significantly different from other previously single or region-based CADx methods.

Chapters 4, and 5 are mainly about feature optimization. Specifically, in chapter 4, I proposed and tested a new approach to develop a quantitative feature analysis and machine learning scheme to predict short-term breast cancer risk, or the likelihood of women having or developing imaging detectable early breast cancer in the next subsequent mammography screening. The results of this study demonstrated that utilizing the LPP-based feature regeneration approach enabled us to create a smaller or compact new feature vector, and yield higher prediction performance than using either all initial image features or a set of selected highly performed features. Besides, this study showed that applying LPP reduces the dimension of feature space and reorganizes the new feature vector to achieve a lower amount of redundancy and maximum variance. Hence, the LPP-regenerated feature vector represents an optimal combination of the highly effective parts of all input features.

In chapter 5, I investigated the feasibility of applying a new approach based on the random projection algorithm (RPA) to generate the optimal feature vectors for training ML models implemented in the CAD schemes of mammograms to classify malignant and benign breast lesions. Study results indicate that applying this RPA approach creates a more compact feature space that can reduce feature correlation or redundancy and provide a subspace with better contrast for the classifier. By comparing with the other three popular feature dimensionality reduction methods in this study, the results also demonstrate that using RPA enables to generate an optimal feature vector to build a machine learning model, which yields significantly higher classification performance. In addition, since creating an optimal feature vector is an essential precondition of building optimal machine learning models, the new method demonstrated in this study is not only limited to CAD schemes of mammograms, it can also be adopted and used by researchers to develop and optimize CAD schemes of other types of medical images to detect and diagnose different types of cancers or diseases in the future.

Finally, chapter 6 is mainly about developing CAD schemes based on DL systems and the challenges that we have with them in medical imaging, how it is possible to deal with them, and the possible solutions to some of these challenges. Thus, I proposed and investigated several new approaches to develop a transfer deep learning CNN model to detect and classify COVID-19 cases using chest X-ray images. Study results demonstrate the value of performing the image pre-processing phase to generate better input image data to build deep learning models. This phase includes removing irrelevant regions, normalizing the image contrast-to-noise ratio, and generating pseudo color images to feed into all three channels of the CNN models in applying the transfer learning method. The reported high classification performance is also promising, which provides a solid foundation to further optimize the deep learning model to detect COVID-19 cases and validate its performance and robustness using large and diverse image datasets in future studies.

The effort and contribution of my research during my Ph.D. study in developing ML/DL-based CAD schemes based on new image processing techniques, feature extraction/optimization algorithms, and machine learning systems resulted in the following list of publications in which I collaborate as a coauthor in them. The proposed CAD schemes for different purposes in my dissertation have been evaluated on large and diverse datasets which their promising results are reported in my publications. Thus, my research and study demonstrate that it would be applicable to utilize my proposed CAD schemes for different medical imaging disease diagnoses and assessments to help radiologists as a second decision-making application to provide more reliable diagnoses about diseases.

7.1.1 Journal Papers:

1. **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. (International Journal of Medical Informatics 2020; 144, 104284).

2. **Heidari M,** Mirniaharikandehei S, Liu W, Hollingsworth AB, Liu H, Zheng B, Development and assessment of a new global mammographic image feature analysis scheme to predict likelihood of malignant cases, IEEE Transactions on Medical Imaging 2020; 39:1235-1244.
3. **Heidari M,** Khuzani AZ, Hollingsworth AB, Danala G, Mirniaharikandehei S, Qiu Y, Liu H, Zheng B, Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm, Physics in Medicine and Biology, 2018; 63:035020.
4. **Heidari M**, Lakshmivarahan S, Mirniaharikandehei S, Danala G, Maryada SK, Liu H, Zheng B, Applying a random projection algorithm to optimize machine learning model for breast lesion classification, IEEE Transactions on Biomedical Engineering, 2021.
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, Computer methods and programs biomedicine, 2021.
6. Danala G, **Heidari M**, Ray B, Desai M, Zheng B, Developing New Quantitative CT Image Markers to Predict Prognosis of Acute Ischemic Stroke Patients, (submitted to Journal Computerized Medical Imaging and Graphics, December 2020).
7. Zheng B, Qiu Y, Aghaei F, Mirniaharikandehei S, **Heidari M**, Danala G. (2019). Developing global image feature analysis models to predict cancer risk and prognosis, Visual Computing for Industry Biomedicine and Art 2019; 2:1-14.
8. Mirniaharikandehei S, VanOsdol J, **Heidari M**, Danala G, Sethuraman SN, Ranjan A, Zheng B, Developing a quantitative ultrasound image feature analysis scheme to assess tumor treatment efficacy using a mouse model, Scientific Reports, 2018; 9: 1-10.

9. Mirniaharikandehei S, Hollingsworth AB, Patel B, **Heidari M**, Liu H, Zheng B, Applying a new computer-aided detection scheme generated imaging marker to predict short-term breast cancer risk, Physics in Medicine & Biology 2018; 63: 105005.
10. Danala G, Patel B, Aghaei F, **Heidari M**, Li J, Wu T, Zheng B, Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms, Annals of biomedical engineering, 2018; 46: 1419-1431.
11. Zargari A, Du Y, **Heidari M**, Thai TC, Gunderson CC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y, Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker, Physics in Medicine & Biology, 2018; 63: 155020.

7.1.2 Conference Proceeding Papers:

1. **Heidari M**, Khuzani AZ, Mirniaharikandehei S, Danala G, Zheng B, Detecting COVID-19 infected pneumonia from X-ray images using a deep learning model with image preprocessing algorithm, Accepted in Proc SPIE 2021.
2. **Heidari M**, Mirniaharikandehei S, Danala G, Zheng B, An optimal machine learning model for breast lesion classification based on random projection algorithm for feature optimization, Accepted in Proc SPIE 2021.
3. **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, Proc SPIE 2020; 11318: 1131816.
4. **Heidari M**, Mirniaharikandehei S, Khuzani AZ, Qian W, Qiu Y, Zheng B, Assessment of a quantitative mammographic imaging marker for breast cancer risk prediction, Proc SPIE 2019; 10952: 109520X.

5. **Heidari M**, Hollingsworth A, Mirniaharikandehei S, Danala G, Qiu Y, Liu H, Zheng B, Assessment of short-term breast cancer risk using a frequency domain correlation based imaging marker, Proc SPIE 2019; 10954: 109541F.
6. **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, Proc SPIE, 2018; 10579: 105790T. Best paper of the conference**.**
7. **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, Proc SPIE, 2018; 10575: 105750P.
8. Danala G, Ray B, **Heidari M**, Desai M, Zheng B, Applying quantitative image markers to predict clinical measures after aneurysmal subarachnoid hemorrhage, Accepted in Proc SPIE 2021.
9. Danala G, **Heidari M**, Mirniaharikandehei S, Desai M, Zheng B, Computer-aided prediction of clinical outcome of treating acute ischemic stroke patients, Accepted in Proc SPIE 2021.
10. Mirniaharikandehei S, **Heidari M**, Danala G, Liu H, Zheng B, A novel feature reduction method to improve performance of machine learning model, Accepted in Proc SPIE 2021.
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, Proc SPIE 2020; 11317: 113171V.
12. Mirniaharikandehei S, **Heidari M**, Danala G, Qian W, Qiu Y, Zheng B, Association of computer-aided detection results and breast cancer risk, Proc SPIE 2019; 10950: 109500I.
13. Zarafshani A, Wang Y, Mirniaharikandehei S, **Heidari M**, Aghaei F, Wang S, Xiang L, Zheng B, Design, fabrication and evaluation of non-imaging, label-free pre-screening tool using quantified bio-electrical tissue profile, Proc SPIE 2019; 10953: 1095304.

14. Danala G, **Heidari M**, Aghaei F, Ray B, Zheng B, Developing a computer-aided image analysis and visualization tool to predict region-specific brain tissue “at risk” for developing acute ischemic stroke, Proc SPIE 2019; 10953: 109530M.
15. Danala G, Aghaei F, **Heidari M**, Wu T, Patel B, Zheng B, Computer-aided classification of breast masses using contrast-enhanced digital mammograms, Proc SPIE 2018; 10575: 105752K.
16. Khuzani AZ, Danala G, **Heidari M**, Du Y, Mashhadi N, Qiu Y, Zheng B, Applying a new unequally weighted feature fusion method to improve CAD performance of classifying breast lesions, Proc SPIE 2018; 10575: 105752L.
17. Wang Y, **Heidari M**, Mirniaharikandehei S, Gong J, Qian W, Qiu Y, Zheng B, A hybrid deep learning approach to predict malignancy of breast lesions using mammograms, Proc SPIE 2018; 10579: 105790V.
18. Mirniaharikandehei S, Zarafshani A, **Heidari M**, Wang Y, Aghaei F, Zheng B, Applying a CAD-generated imaging marker to assess short-term breast cancer risk, Proc SPIE 2018; 10575: 105753F.

7.2 Future works

Despite extensive research efforts and progress made in the CAD field, researchers still face many challenges in developing CAD schemes for clinical applications. For example, in developing CAD schemes based on classical machine learning algorithms, which use image features to train classification models to predict the likelihood of the analyzed regions depicting or patterns representing diseases, due to the great heterogeneity of disease patterns and the limited size of image datasets, it is hard to have perfect segmentation algorithms for lesion or ROI selection. Also, modeling the human vision system of radiologists in computer vision algorithms to extract meaningful features is difficult. Furthermore, identifying a small and optimal image feature vector to build highly performed and robust machine learning models

remains difficult. It means although the already available CAD schemes present high performance, they still need to go for further analyses either in clinical practices or on large global and diverse datasets provided from different sites.

Based on my experience I have got during my Ph.D. study, I will continue my effort to explore more general techniques on each of these challenges in the next steps of my research life. For instance, since in recent developments, deep learning technology based on convolutional neural networks becomes the state of art of artificial intelligence applying in medical imaging for all the areas covered by classical machine learning schemes. Also, deep learning schemes have shown many promising results in other research fields for the segmentation task. Thus, a perfect segmentation algorithm as the initial phase of any ML/DL based CAD scheme would be a solution to many available optimization problems. It would be helpful if we apply them to the mammograms to remove chest walls, breast surface, nipple, and all the other artifacts perfectly and then apply global feature extraction algorithms on the breast's interior space. The combination of deep learning systems to do the segmentation part, and classical machine learning-based systems to do the classification part of any CAD scheme may help better to model human vision system and present more acceptable CAD schemes to radiologists. Besides, as we presented in chapter six, we can take advantage of DL schemes to either directly develop a new complete DL- based classification scheme from scratch or use the deep learning model as a feature extractor to replace the handcrafted features in the conventional machine learning approach. The deep learning model can act as a feature generator, and then robust extracted features are serviceable for any classical machine learning scheme like SVM.

Besides the research and studies I have done so far, I propose doing additional tasks with DL-based techniques like lesion classification or lesion localization as another task hard to get with classical ML-based algorithms. DL-based CAD schemes are usually more accurate in comparison with the available classical ML-based CAD schemes. I propose to develop and

evaluate a new CAD scheme and decision-making supporting tool for radiologists, which will be developed based on CNNs. I will then figure out if it is possible to combine it with classical data optimization methods, I proposed in my other research for classical ML-based models (chapter 4,5). By combining my previous works and additional new work in my Ph.D. researches and studies, I can comprehensively investigate different machine learning algorithms' challenges as a frame of CAD systems and develop more robust algorithms. Since the proposed tasks are in the area of deep learning, which is the state of the art of research in the medical imaging field, I believe my research's success will provide a significant contribution to the development of CAD systems in the future of 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] R. M. Nishikawa and D. Gur, "CADe for early detection of breast cancer—current status and why we need to continue to explore new approaches," *Academic radiology,* vol. 21, no. 10, pp. 1320-1321, 2014.

[2] Z. Huo, M. L. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: effectiveness of computer-aided diagnosis—observer study with independent database of mammograms," *Radiology,* vol. 224, no. 2, pp. 560-568, 2002.

[3] J. J. Fenton *et al.*, "Influence of computer-aided detection on performance of screening mammography," *New England Journal of Medicine,* vol. 356, no. 14, pp. 1399-1409, 2007.

[4] J. Brodersen and V. D. Siersma, "Long-term psychosocial consequences of false-positive screening mammography," *The Annals of Family Medicine,* vol. 11, no. 2, pp. 106-115, 2013.

[5] M. J. Yaffe and J. G. Mainprize, "Risk of radiation-induced breast cancer from mammographic screening," *Radiology,* vol. 258, no. 1, pp. 98-105, 2011.

[6] D. S. Buist *et al.*, "Influence of annual interpretive volume on screening mammography performance in the United States," *Radiology,* vol. 259, no. 1, pp. 72-84, 2011.

[7] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics,* vol. 31, no. 4-5, pp. 198-211, 2007.

[8] Y. Wang, F. Aghaei, A. Zarafshani, Y. Qiu, W. Qian, and B. Zheng, "Computer-aided classification of mammographic masses using visually sensitive image features," *Journal of X-ray Science and technology,* vol. 25, no. 1, pp. 171-186, 2017.

[9] P. Lambin *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer,* vol. 48, no. 4, pp. 441-446, 2012.

[10] G. Danala *et al.*, "Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms," *Annals of biomedical engineering,* vol. 46, no. 9, pp. 1419-1431, 2018.

[11] M. Hatt, C. Parmar, J. Qi, and I. El Naqa, "Machine (deep) learning methods for image processing and radiomics," *IEEE Transactions on Radiation and Plasma Medical Sciences,* vol. 3, no. 2, pp. 104-108, 2019.

[12] Y. Qiu *et al.*, "A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology," *Journal of X-ray Science and Technology,* vol. 25, no. 5, pp. 751-763, 2017.

[13] F. Gao *et al.*, "SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis," *Computerized Medical Imaging and Graphics,* vol. 70, pp. 53-62, 2018.

[14] X. Liu, L. Ma, L. Song, Y. Zhao, X. Zhao, and C. Zhou, "Recognizing common CT imaging signs of lung diseases through a new feature selection method based on Fisher criterion and genetic optimization," *IEEE J. Biomed. Health Informatics,* vol. 19, no. 2, pp. 635-647, 2015.

[15] Z. Zhang and E. Sejdić, "Radiological images and machine learning: trends, perspectives, and prospects," *Computers in biology and medicine,* vol. 108, pp. 354-370, 2019.

[16] M. M. Dundar, G. Fung, B. Krishnapuram, and R. B. Rao, "Multiple-instance learning algorithms for computer-aided detection," *IEEE Transactions on Biomedical Engineering,* vol. 55, no. 3, pp. 1015-1021, 2008.

[17] R. M. Summers, "Improving the accuracy of CTC interpretation: computer-aided detection," *Gastrointestinal Endoscopy Clinics,* vol. 20, no. 2, pp. 245-257, 2010.

[18] Y. Li, M. Fan, H. Cheng, P. Zhang, B. Zheng, and L. Li, "Assessment of global and local region-based bilateral mammographic feature asymmetry to predict short-term breast cancer risk," *Physics in Medicine & Biology,* vol. 63, no. 2, p. 025004, 2018.

[19] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine & Biology,* vol. 58, no. 13, p. R97, 2013.

[20] S. Mirniaharikandehei *et al.*, "Developing a quantitative ultrasound image feature analysis scheme to assess tumor treatment efficacy using a mouse model," *Scientific reports,* vol. 9, no. 1, pp. 1-10, 2019.

[21] Y. Kusumawardani, P. Prajitno, and D. Soejoko, "Computer-Aided Diagnosis (CAD) to Detect Brain Abnormality from PET Image using Artificial Neural Network (ANN)," in *Journal of Physics: Conference Series*, 2020, vol. 1505, no. 1: IOP Publishing, p. 012002.

[22] Y. Yan, E. Somer, and V. Grau, "Classification of amyloid PET images using novel features for early diagnosis of Alzheimer's disease and mild cognitive impairment conversion," *Nuclear medicine communications,* vol. 40, no. 3, pp. 242-248, 2019.

[23] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiology of aging,* vol. 29, no. 4, pp. 514-523, 2008.

[24] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer informatics,* vol. 2, p. 117693510600200030, 2006.

[25] G. Currie, K. E. Hawk, E. Rohren, A. Vial, and R. Klein, "Machine learning and deep learning in medical imaging: intelligent imaging," *Journal of medical imaging and radiation sciences,* vol. 50, no. 4, pp. 477-487, 2019.

[26] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[27] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics,* vol. 37, no. 2, pp. 505-515, 2017.

[28] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik,* vol. 29, no. 2, pp. 102-127, 2019.

[29] S. Leger *et al.*, "A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling," *Scientific reports,* vol. 7, no. 1, pp. 1-11, 2017.

[30] T. M. Deist *et al.*, "Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers," *Medical physics,* vol. 45, no. 7, pp. 3449-3459, 2018.

[31] P. Flach, *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.

[32] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.

[33] S. J. Roberts, D. Husmeier, I. Rezek, and W. Penny, "Bayesian approaches to Gaussian mixture modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1133-1142, 1998.

[34] S. C. Park *et al.*, "Computer-aided detection of early interstitial lung diseases using low-dose CT images," *Physics in Medicine & Biology,* vol. 56, no. 4, p. 1139, 2011.

[35] S. Li *et al.*, "A novel approach of computer-aided detection of focal ground-glass opacity in 2D lung CT images," in *Medical Imaging 2013: Computer-Aided Diagnosis*, 2013, vol. 8670: International Society for Optics and Photonics, p. 86702W.

[36] R. Shen, I. Cheng, and A. Basu, "A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs," *IEEE transactions on biomedical engineering,* vol. 57, no. 11, pp. 2646-2656, 2010.

[37] M. Kuhn and K. Johnson, "An introduction to feature selection," in *Applied predictive modeling*: Springer, 2013, pp. 487-519.

[38] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering,* vol. 40, no. 1, pp. 16-28, 2014.

[39] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Histogram-based features selection and volume of interest ranking for brain PET image classification," *IEEE journal of translational engineering in health and medicine,* vol. 6, pp. 1-12, 2018.

[40] P. Morgado, M. Silveira, and J. S. Marques, "Diagnosis of Alzheimer's disease using 3D local binary patterns," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 1, no. 1, pp. 2-12, 2013.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems,* vol. 25, pp. 1097-1105, 2012.

[42] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology,* vol. 10, no. 3, pp. 257-273, 2017.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[45] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.

[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[47] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697-8710.

[48] M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology,* vol. 15, no. 3, pp. 512-520, 2018.

[49] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 07, pp. 13001-13008.

[50] H. R. Roth *et al.*, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1170-1181, 2015.

[51] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1207-1216, 2016.

[52] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, J. Wei, and K. Cha, "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Medical physics,* vol. 43, no. 12, pp. 6654-6666, 2016.

[53] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Medical physics,* vol. 44, no. 10, pp. 5162-5171, 2017.

[54] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging,* vol. 3, no. 3, p. 034501, 2016.

[55] G. Choy *et al.*, "Current applications and future impact of machine learning in radiology," *Radiology,* vol. 288, no. 2, pp. 318-328, 2018.

[56] M. Kohli, L. M. Prevedello, R. W. Filice, and J. R. Geis, "Implementing machine learning in radiology practice and research," *American journal of roentgenology,* vol. 208, no. 4, pp. 754-760, 2017.

[57] J. E. Burns, J. Yao, H. Muñoz, and R. M. Summers, "Automated detection, localization, and classification of traumatic vertebral body fractures in the thoracic and lumbar spine at CT," *Radiology,* vol. 278, no. 1, pp. 64-73, 2016.

[58] A. A. A. Setio *et al.*, "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1160-1169, 2016.

[59] M. H. Le *et al.*, "Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks," *Physics in Medicine & Biology,* vol. 62, no. 16, p. 6497, 2017.

[60] H. Lee *et al.*, "Fully automated deep learning system for bone age assessment," *Journal of digital imaging,* vol. 30, no. 4, pp. 427-441, 2017.

[61] P. Gu, W.-M. Lee, M. A. Roubidoux, J. Yuan, X. Wang, and P. L. Carson, "Automated 3D ultrasound image segmentation to aid breast cancer image interpretation," *Ultrasonics,* vol. 65, pp. 51-58, 2016.

[62] S. Kido, Y. Hirano, and S. Mabu, "Deep learning for pulmonary image analysis: classification, detection, and segmentation," *Deep Learning in Medical Image Analysis,* pp. 47-58, 2020.

[63] E. B. Cole, Z. Zhang, H. S. Marques, R. Edward Hendrick, M. J. Yaffe, and E. D. Pisano, "Impact of computer-aided detection systems on radiologist accuracy with digital mammography," *American Journal of Roentgenology,* vol. 203, no. 4, pp. 909-916, 2014.

[64] S. G. Orel and M. D. Schnall, "MR imaging of the breast for the detection, diagnosis, and staging of breast cancer," *Radiology,* vol. 220, no. 1, pp. 13-30, 2001.

[65] A. Mansoor *et al.*, "Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends," *Radiographics,* vol. 35, no. 4, pp. 1056-1076, 2015.

[66] K. Doman, T. Konishi, and Y. Mekada, "Lesion Image Synthesis Using DCGANs for Metastatic Liver Cancer Detection," *Deep Learning in Medical Image Analysis,* pp. 95-106, 2020.

[67] X. Zhou, "Automatic Segmentation of Multiple Organs on 3D CT Images by Using Deep Learning Approaches," *Deep Learning in Medical Image Analysis,* pp. 135-147, 2020.

[68] A. P. Yow *et al.*, "Techniques and Applications in Skin OCT Analysis," *Deep Learning in Medical Image Analysis,* pp. 149-163, 2020.

[69] Z. Liao *et al.*, "Echocardiography view classification using quality transfer star generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019: Springer, pp. 687-695.

[70] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *International conference on medical image computing and computer-assisted intervention*, 2017: Springer, pp. 417-425.

[71] S. Ghose *et al.*, "A review of segmentation and deformable registration methods applied to adaptive cervical cancer radiation therapy treatment planning," *Artificial intelligence in medicine,* vol. 64, no. 2, pp. 75-87, 2015.

[72] 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," *Computer methods and programs in biomedicine,* vol. 144, pp. 97-104, 2017.

[73] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: state of the art and future directions," *Journal of digital imaging,* vol. 30, no. 4, pp. 449-459, 2017.

[74] S. G. Sapate, A. Mahajan, S. N. Talbar, N. Sable, S. Desai, and M. Thakur, "Radiomics based detection and characterization of suspicious lesions on full field digital mammograms," *Computer methods and programs in biomedicine,* vol. 163, pp. 1-20, 2018.

[75] H. Li, K. R. Mendel, L. Lan, D. Sheth, and M. L. Giger, "Digital mammography in breast cancer: additive value of radiomics of breast parenchyma," *Radiology,* vol. 291, no. 1, pp. 15-20, 2019.

[76] 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 transactions on medical imaging,* vol. 35, no. 7, pp. 1719-1728, 2016.

[77] M. Tan, J. Pu, S. Cheng, H. Liu, and B. Zheng, "Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk," *Annals of Biomedical Engineering,* vol. 43, no. 10, pp. 2416-2428, 2015.

[78] 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," *Computer methods and programs in biomedicine,* vol. 179, p. 104995, 2019.

[79] Q. Yang, L. Li, J. Zhang, G. Shao, and B. Zheng, "A computerized global MR image feature analysis scheme to assist diagnosis of breast cancer: a preliminary assessment," *European journal of radiology,* vol. 83, no. 7, pp. 1086-1091, 2014.

[80] P. Casti *et al.*, "Estimation of the breast skin-line in mammograms using multidirectional Gabor filters," *Computers in Biology and Medicine,* vol. 43, no. 11, pp. 1870-1881, 2013.

[81] P. Casti *et al.*, "Automatic detection of the nipple in screen-film and full-field digital mammograms using a novel Hessian-based method," *Journal of digital imaging,* vol. 26, no. 5, pp. 948-957, 2013.

[82] L. Sellami, O. B. Sassi, and A. B. Hamida, "Breast cancer ultrasound images' sequence exploration using BI-RADS features' extraction: towards an advanced clinical aided tool for precise lesion characterization," *IEEE transactions on nanobioscience,* vol. 14, no. 7, pp. 740-745, 2015.

[83] M. Heidari, S. Mirniaharikandehei, W. Liu, A. B. Hollingsworth, H. Liu, and B. Zheng, "Development and assessment of a new global mammographic image feature analysis scheme to predict likelihood of malignant cases," *IEEE Transactions on Medical Imaging,* vol. 39, no. 4, pp. 1235-1244, 2019.

[84] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer," *Neurocomputing,* vol. 154, pp. 1-14, 2015.

[85] I. Buciu and A. Gacsadi, "Directional features for automatic tumor classification of mammogram images," *Biomedical Signal Processing and Control,* vol. 6, no. 4, pp. 370-378, 2011.

[86] J. A. Putra, "Mammogram classification scheme using 2D-discrete wavelet and local binary pattern for detection of breast cancer," in *Journal of Physics: Conference Series*, 2018, vol. 1008, no. 1: IOP Publishing, p. 012004.

[87] X. Long and C. Wyatt, "An automatic unsupervised classification of MR images in Alzheimer's disease," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010: IEEE, pp. 2910-2917.

[88] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.

[89] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning,* vol. 46, no. 1-3, pp. 389-422, 2002.

[90] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 27, no. 8, pp. 1226-1238, 2005.

[91] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine learning,* vol. 53, no. 1-2, pp. 23-69, 2003.

[92] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval*, 2007, pp. 197-206.

[93] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering,* vol. 19, pp. 221-248, 2017.

[94] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153-160.

[95] M. Heidari *et al.*, "Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm," *Physics in Medicine & Biology,* vol. 63, no. 3, p. 035020, 2018.

[96] M. Heidari *et al.*, "Applying a random projection algorithm to optimize machine learning model for breast lesion classification," *IEEE Transactions on Biomedical Engineering,* 2021.

[97] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE transactions on neural networks and learning systems,* vol. 30, no. 5, pp. 1581-1586, 2018.

[98] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition,* vol. 43, no. 1, pp. 331-341, 2010.

[99] Y. Gao, X. Shan, Z. Hu, D. Wang, Y. Li, and X. Tian, "Extended compressed tracking via random projection based on MSERs and online LS-SVM learning," *Pattern Recognition,* vol. 59, pp. 245-254, 2016.

[100] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, "Fast indoor scene description for blind people with multiresolution random projections," *Journal of Visual Communication and Image Representation,* vol. 44, pp. 95-105, 2017.

[101] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE transactions on neural networks and learning systems,* vol. 27, no. 4, pp. 809-821, 2015.

[102] L. Lu, L. Yapeng, and Z. Hongyuan, "Benign and malignant solitary pulmonary nodules classification based on CNN and SVM," in *Proceedings of the International Conference on Machine Vision and Applications*, 2018, pp. 46-50.

[103] M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Three-class mammogram classification based on descriptive CNN features," *BioMed research international,* vol. 2017, 2017.

[104] Z. A. Sejuti and M. S. Islam, "An Efficient Method to Classify Brain Tumor using CNN and SVM," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021: IEEE, pp. 644-648.

[105] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE journal of biomedical and health informatics,* vol. 21, no. 1, pp. 31-40, 2016.

[106] A. H. Shayma'a, M. S. Sayed, M. I. Abdalla, and M. A. Rashwan, "Breast cancer masses classification using deep convolutional neural networks and transfer learning," *Multimedia Tools and Applications,* vol. 79, no. 41, pp. 30735-30768, 2020.

[107] R. Arora, P. K. Rai, and B. Raman, "Deep feature–based automatic classification of mammograms," *Medical & biological engineering & computing,* pp. 1-13, 2020.

[108] M. A. Al-Antari, M. A. Al-Masni, M.-T. Choi, S.-M. Han, and T.-S. Kim, "A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification," *International journal of medical informatics,* vol. 117, pp. 44-54, 2018.

[109] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[110] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[111] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms," *International journal of medical informatics,* vol. 144, p. 104284, 2020.

[112] T. Force, "Breast Cancer Screening: A Review of Current Guidelines."

[113] M. Heidari *et al.*, "An optimal machine learning model for breast lesion classification based on random projection algorithm for feature optimization," in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, 2021, vol. 11601: International Society for Optics and Photonics, p. 116010L.

[114] K. M. Kelly, J. Dean, W. S. Comulada, and S.-J. Lee, "Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts," *European radiology,* vol. 20, no. 3, pp. 734-742, 2010.

[115] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, W. Qian, Y. Qiu, and B. Zheng, "Assessment of a quantitative mammographic imaging marker for breast cancer risk prediction," in *Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment*, 2019, vol. 10952: International Society for Optics and Photonics, p. 109520X.

[116] D. Woodard, A. Gelfand, W. Barlow, and J. Elmore, "Performance assessment for radiologists interpreting screening mammography," *Statistics in medicine,* vol. 26, no. 7, pp. 1532-1551, 2007.

[117] R. A. Hubbard, K. Kerlikowske, C. I. Flowers, B. C. Yankaskas, W. Zhu, and D. L. Miglioretti, "Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study," *Annals of internal medicine,* vol. 155, no. 8, pp. 481-492, 2011.

[118] M. Heidari, A. Z. Khuzani, G. Danala, S. Mirniaharikandehei, W. Qian, and B. Zheng, "Applying a machine learning model using a locally preserving projection based feature regeneration algorithm to predict breast cancer risk," in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, 2018, vol. 10579: International Society for Optics and Photonics, p. 105790T.

[119] M. Heidari, S. Mirniaharikandehei, G. Danala, Y. Qiu, and B. Zheng, "A new case-based CAD scheme using a hierarchical SSIM feature extraction method to classify between malignant and benign cases," in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, 2020, vol. 11318: International Society for Optics and Photonics, p. 1131816.

[120] M. Tan, W. Qian, J. Pu, H. Liu, and B. Zheng, "A new approach to develop computer-aided detection schemes of digital mammograms," *Physics in Medicine & Biology,* vol. 60, no. 11, p. 4413, 2015.

[121] 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," *Journal of Magnetic Resonance Imaging,* vol. 44, no. 5, pp. 1099-1106, 2016.

[122] Y. Wang *et al.*, "Quantitative measurement of adiposity using CT images to predict the benefit of bevacizumab-based chemotherapy in epithelial ovarian cancer patients," *Oncology letters,* vol. 12, no. 1, pp. 680-686, 2016.

[123] B. Zheng *et al.*, "Multiview-based computer-aided detection scheme for breast masses," *Medical physics,* vol. 33, no. 9, pp. 3135-3143, 2006.

[124] 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," *The British journal of radiology,* vol. 85, no. 1014, pp. e153-e161, 2012.

[125] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing,* vol. 13, no. 4, pp. 600-612, 2004.

[126] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: A new image similarity index," *IEEE transactions on image processing,* vol. 18, no. 11, pp. 2385-2401, 2009.

[127] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications*. Academic press, 2014.

[128] C. S. Burrus, *Fast fourier transforms*. Lulu. com, 2012.

[129] X. Fan, Y. Lu, and W. Gao, "A novel coefficient scanning scheme for directional spatial prediction-based image compression," in *2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No. 03TH8698)*, 2003, vol. 2: IEEE, pp. II-557.

[130] R. Rouhi and M. Jafari, "Classification of benign and malignant breast tumors based on hybrid level set segmentation," *Expert Systems with Applications,* vol. 46, pp. 45-59, 2016.

[131] K. Yin, S. Yan, C. Song, and B. Zheng, "A robust method for segmenting pectoral muscle in mediolateral oblique (MLO) mammograms," *International journal of computer assisted radiology and surgery,* vol. 14, no. 2, pp. 237-248, 2019.

[132] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Deepmatching: Hierarchical deformable dense matching," *International Journal of Computer Vision,* vol. 120, no. 3, pp. 300-323, 2016.

[133] M. Heidari, S. Samavi, S. M. R. Soroushmehr, S. Shirani, N. Karimi, and K. Najarian, "Framework for robust blind image watermarking based on classification of attacks," *Multimedia Tools and Applications,* vol. 76, no. 22, pp. 23459-23479, 2017.

[134] Y. Wang and P. Moulin, "Optimized feature extraction for learning-based image steganalysis," *IEEE Transactions on Information Forensics and Security,* vol. 2, no. 1, pp. 31-45, 2007.

[135] D. Lederman, B. Zheng, X. Wang, X. H. Wang, and D. Gur, "Improving breast cancer risk stratification using resonance-frequency electrical impedance spectroscopy through fusion of multiple classifiers," *Annals of biomedical engineering,* vol. 39, no. 3, pp. 931-945, 2011.

[136] M. Tan, J. Pu, and B. Zheng, "Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model," *International journal of computer assisted radiology and surgery,* vol. 9, no. 6, pp. 1005-1020, 2014.

[137] A. Zargari *et al.*, "Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker," *Physics in Medicine & Biology,* vol. 63, no. 15, p. 155020, 2018.

[138] 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," *Physics in Medicine & Biology,* vol. 63, no. 10, p. 105005, 2018.

[139] S. Narod, P. Sun, C. Wall, C. Baines, and A. Miller, "Impact of screening mammography on mortality from breast cancer before age 60 in women 40 to 49 years of age," *Current Oncology,* vol. 21, no. 5, p. 217, 2014.

[140] D. Saslow *et al.*, "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography," *CA: a cancer journal for clinicians,* vol. 57, no. 2, pp. 75-89, 2007.

[141] L. Berlin and F. M. Hall, "More mammography muddle: emotions, politics, science, costs, and polarization," *Radiology,* vol. 255, no. 2, pp. 311-316, 2010.

[142] P. A. Carney *et al.*, "Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography," *Annals of internal medicine,* vol. 138, no. 3, pp. 168-175, 2003.

[143] 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," *Visual Computing for Industry, Biomedicine, and Art,* vol. 2, no. 1, pp. 1-14, 2019.

[144] O. W. Brawley, "Risk-based mammography screening: an effort to maximize the benefits and minimize the harms," *Annals of internal medicine,* vol. 156, no. 9, pp. 662-663, 2012.

[145] S. Mirniaharikandehei, M. Heidari, G. Danala, S. Lakshmivarahan, and B. Zheng, "A novel feature reduction method to improve performance of machine learning model," in *Medical Imaging 2021: Computer-Aided Diagnosis*, 2021, vol. 11597: International Society for Optics and Photonics, p. 1159726.

[146] B. Zheng, J. H. Sumkin, M. L. Zuley, X. Wang, A. H. Klym, and D. Gur, "Bilateral mammographic density asymmetry and breast cancer risk: a preliminary assessment," *European journal of radiology,* vol. 81, no. 11, pp. 3222-3228, 2012.

[147] M. H. Gail *et al.*, "Projecting individualized probabilities of developing breast cancer for white females who are being examined annually," *JNCI: Journal of the National Cancer Institute,* vol. 81, no. 24, pp. 1879-1886, 1989.

[148] J. Tyrer, S. W. Duffy, and J. Cuzick, "A breast cancer prediction model incorporating familial and personal risk factors," *Statistics in medicine,* vol. 23, no. 7, pp. 1111-1130, 2004.

[149] M. H. Gail and P. L. Mai, "Comparing breast cancer risk assessment models," ed: Oxford University Press, 2010.

[150] M. van Zitteren, J. B. van der Net, S. Kundu, A. N. Freedman, C. M. van Duijn, and A. C. J. Janssens, "Genome-based prediction of breast cancer risk in the general population: a modeling study based on meta-analyses of genetic associations," *Cancer Epidemiology and Prevention Biomarkers,* vol. 20, no. 1, pp. 9-22, 2011.

[151] J. Wei *et al.*, "Association of computerized mammographic parenchymal pattern measure with breast cancer risk: a pilot case-control study," *Radiology,* vol. 260, no. 1, pp. 42-49, 2011.

[152] M. Heidari *et al.*, "Assessment of short-term breast cancer risk using a frequency domain correlation based imaging marker," in *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, 2019, vol. 10954: International Society for Optics and Photonics, p. 109541F.

[153] E. Amir, O. C. Freedman, B. Seruga, and D. G. Evans, "Assessing women at high risk of breast cancer: a review of risk assessment models," *JNCI: Journal of the National Cancer Institute,* vol. 102, no. 10, pp. 680-691, 2010.

[154] W. A. Berg, C. Campassi, P. Langenberg, and M. J. Sexton, "Breast Imaging Reporting and Data System: inter-and intraobserver variability in feature analysis and final assessment," *American Journal of Roentgenology,* vol. 174, no. 6, pp. 1769-1777, 2000.

[155] C. N. Damases, P. C. Brennan, C. Mello-Thoms, and M. F. McEntee, "Mammographic breast density assessment using automated volumetric software and breast imaging reporting and data system (BIRADS) categorization by expert radiologists," *Academic radiology,* vol. 23, no. 1, pp. 70-77, 2016.

[156] D. B. Kopans, "Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk," *Radiology,* vol. 246, no. 2, pp. 348-353, 2008.

[157] B. Zheng, M. Tan, P. Ramalingam, and D. Gur, "Association between Computed Tissue Density Asymmetry in Bilateral Mammograms and Near-term Breast Cancer Risk," *The breast journal,* vol. 20, no. 3, pp. 249-257, 2014.

[158] B. Zheng *et al.*, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," *Medical physics,* vol. 33, no. 1, pp. 111-117, 2006.

[159] 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," *Medical engineering & physics,* vol. 33, no. 8, pp. 934-942, 2011.

[160] B. Zheng, Y.-H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Academic radiology,* vol. 2, no. 11, pp. 959-966, 1995.

[161] T. Mitchell, "Machine learning (WBC/McGraw-Hill, Boston)," ed: MA, 1997.

[162] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research,* vol. 10, no. 2, 2009.

[163] Y. Leng, X. Xu, and G. Qi, "Combining active learning and semi-supervised learning to construct SVM classifier," *Knowledge-Based Systems,* vol. 44, pp. 121-131, 2013.

[164] Q. Li and K. Doi, "Reduction of bias and variance for evaluation of computer-aided diagnostic schemes," *Medical physics,* vol. 33, no. 4, pp. 868-875, 2006.

[165] S. Yan, W. Qian, Y. Guan, and B. Zheng, "Improving lung cancer prognosis assessment by incorporating synthetic minority oversampling technique and score fusion method," *Medical Physics,* vol. 43, no. 6Part1, pp. 2694-2703, 2016.

[166] S. C. Park, R. Sukthankar, L. Mummert, M. Satyanarayanan, and B. Zheng, "Optimization of reference library used in content-based medical image retrieval scheme," *Medical Physics,* vol. 34, no. 11, pp. 4331-4339, 2007.

[167] J. Katzen and K. Dodelzon, "A review of computer aided detection in mammography," *Clinical imaging,* vol. 52, pp. 305-309, 2018.

[168] A. Z. Khuzani *et al.*, "Applying a new unequally weighted feature fusion method to improve CAD performance of classifying breast lesions," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, vol. 10575: International Society for Optics and Photonics, p. 105752L.

[169] S. Mirniaharikandehei, A. Zarafshani, M. Heidari, Y. Wang, F. Aghaei, and B. Zheng, "Applying a CAD-generated imaging marker to assess short-term breast cancer risk," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, vol. 10575: International Society for Optics and Photonics, p. 105753F.

[170] M. Heidari, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving performance of breast cancer risk prediction using a new CAD-based region segmentation scheme," in *Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, vol. 10575: International Society for Optics and Photonics, p. 105750P.

[171] J. Yin *et al.*, "A radiomics signature to identify malignant and benign liver tumors on plain CT images," *Journal of X-ray science and technology,* no. Preprint, pp. 1-12, 2020.

[172] Z.-Q. Sun, S.-D. Hu, J. Li, T. Wang, S.-F. Duan, and J. Wang, "Radiomics study for differentiating gastric cancer from gastric stromal tumor based on contrast-enhanced CT images," *Journal of X-ray Science and Technology,* vol. 27, no. 6, pp. 1021-1031, 2019.

[173] A. Z. Khuzani, N. Mashhadi, M. Heidari, D. Khaledyan, and S. Teymoori, "Applying a new feature fusion method to classify breast lesions," in *Medical Imaging 2021: Computer-Aided Diagnosis*, 2021, vol. 11597: International Society for Optics and Photonics, p. 1159711.

[174] X. Wang, L. Li, W. Liu, W. Xu, D. Lederman, and B. Zheng, "An interactive system for computer-aided diagnosis of breast masses," *Journal of digital imaging,* vol. 25, no. 5, pp. 570-579, 2012.

[175] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE transactions on Systems, Man, and Cybernetics,* no. 4, pp. 269-285, 1976.

[176] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics,* no. 6, pp. 610-621, 1973.

[177] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing,* vol. 4, no. 2, pp. 172-179, 1975.

[178] M. Z. Do Nascimento, A. S. Martins, L. A. Neves, R. P. Ramos, E. L. Flores, and G. A. Carrijo, "Classification of masses in mammographic image using wavelet domain features and polynomial classifier," *Expert Systems with Applications,* vol. 40, no. 15, pp. 6213-6221, 2013.

[179] N. R. Mudigonda, R. Rangayyan, and J. L. Desautels, "Gradient and texture analysis for the classification of mammographic masses," *IEEE transactions on medical imaging,* vol. 19, no. 10, pp. 1032-1043, 2000.

[180] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *International conference on database theory*, 2001: Springer, pp. 420-434.

[181] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018.

[182] C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, *Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop, SLSFS 2005 Bohinj, Slovenia, February 23-25, 2005, Revised Selected Papers*. Springer, 2006.

[183] S. Dasgupta and A. Gupta, "An elementary proof of the Johnson-Lindenstrauss lemma," *International Computer Science Institute, Technical Report,* vol. 22, no. 1, pp. 1-5, 1999.

[184] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas, "Random projections for linear support vector machines," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 8, no. 4, pp. 1-25, 2014.

[185] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas, "Random projections for support vector machines," in *Artificial intelligence and statistics*, 2013: PMLR, pp. 498-506.

[186] K. Sridharan, S. Shalev-Shwartz, and N. Srebro, "Fast rates for regularized objectives," *Advances in neural information processing systems,* vol. 21, pp. 1545-1552, 2008.

[187] W. Gómez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast

ultrasound," *IEEE transactions on medical imaging,* vol. 31, no. 10, pp. 1889-1899, 2012.

[188] M. J. Zdilla, S. A. Hatfield, K. A. McLean, L. M. Cyrus, J. M. Laslo, and H. W. Lambert, "Circularity, solidity, axes of a best fit ellipse, aspect ratio, and roundness of the foramen ovale: a morphometric analysis with neurosurgical considerations," *The Journal of craniofacial surgery,* vol. 27, no. 1, p. 222, 2016.

[189] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 287-296.

[190] H. D. Nelson, E. S. O'Meara, K. Kerlikowske, S. Balch, and D. Miglioretti, "Factors associated with rates of false-positive and false-negative results from digital mammography screening: an analysis of registry data," *Annals of internal medicine,* vol. 164, no. 4, pp. 226-235, 2016.

[191] H.-P. Chan, R. K. Samala, and L. M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," *The British journal of radiology,* vol. 93, no. 1108, p. 20190580, 2019.

[192] C. Tao *et al.*, "New one-step model of breast tumor locating based on deep learning," *Journal of X-ray Science and Technology,* vol. 27, no. 5, pp. 839-856, 2019.

[193] P. Lei *et al.*, "Clinical and computed tomographic (CT) images characteristics in the patients with COVID-19 infection: What should radiologists need to know?," *Journal of X-ray science and technology,* no. Preprint, pp. 1-13, 2020.

[194] A. Z. Khuzani, M. Heidari, and S. A. Shariati, "COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images," *medRxiv,* 2020.

[195] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849,* 2020.

[196] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Detecting COVID-19 infected pneumonia from x-ray images using a deep learning model with image preprocessing algorithm," in *Medical Imaging 2021: Computer-Aided Diagnosis*, 2021, vol. 11597: International Society for Optics and Photonics, p. 115970V.

[197] W.-c. Dai *et al.*, "CT imaging and differential diagnosis of COVID-19," *Canadian Association of Radiologists Journal,* vol. 71, no. 2, pp. 195-200, 2020.

[198] M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu, and A. T. Sahlol, "New machine learning method for image-based diagnosis of COVID-19," *PloS one,* vol. 15, no. 6, p. e0235187, 2020.

[199] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv," *Published online April,* vol. 24, p. 2020.02, 2020.

[200] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *medRxiv,* 2020.

[201] P. K. Sethy and S. K. Behera, "Detection of coronavirus disease (covid-19) based on deep features," *Preprints,* vol. 2020030300, p. 2020, 2020.

[202] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine,* p. 1, 2020.

[203] A. I. Khan, J. L. Shah, and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine,* p. 105581, 2020.

[204] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in Medicine Unlocked,* p. 100360, 2020.

[205] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine,* p. 103792, 2020.

[206] L. Wang and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," *arXiv preprint arXiv:2003.09871,* 2020.

[207] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055,* 2020.

[208] D. Kermany, K. Zhang, and M. Goldbaum, "Large dataset of labeled optical coherence tomography (oct) and chest x-ray images," *Mendeley Data, v3 http://dx. doi. org/10.17632/rscbjbr9sj,* vol. 3, 2018.

[209] M. E. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *arXiv preprint arXiv:2003.13145,* 2020.

[210] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," *The Lancet,* vol. 395, no. 10223, pp. 507-513, 2020.

[211] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," *arXiv preprint arXiv:2003.11597,* 2020.

[212] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell,* vol. 172, no. 5, pp. 1122-1131. e9, 2018.

[213] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, no. 10, pp. 1345-1359, 2009.

[214] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision,* vol. 115, no. 3, pp. 211-252, 2015.

[215] M. Pak and S. Kim, "A review of deep learning in image recognition," in *2017 4th international conference on computer applications and information processing technology (CAIPT)*, 2017: IEEE, pp. 1-3.

[216] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *ICML*, 2010.

[217] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[218] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access,* vol. 8, pp. 114822-114832, 2020.

[219] G. King and L. Zeng, "Logistic regression in rare events data," *Political analysis,* vol. 9, no. 2, pp. 137-163, 2001.

[220] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621,* 2017.

[221] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica: Biochemia medica,* vol. 22, no. 3, pp. 276-282, 2012.

[222] X. Zhao *et al.*, "Deep CNN models for pulmonary nodule classification: Model modification, model integration, and transfer learning," *Journal of X-ray Science and Technology,* vol. 27, no. 4, pp. 615-629, 2019.

[223] K. Wang, B. K. Patel, L. Wang, T. Wu, B. Zheng, and J. Li, "A dual-mode deep transfer learning (D2TL) system for breast cancer detection using contrast enhanced digital mammograms," *IISE Transactions on Healthcare Systems Engineering,* vol. 9, no. 4, pp. 357-370, 2019.

[224] U. J. Schoepf and P. Costello, "CT angiography for diagnosis of pulmonary embolism: state of the art," *Radiology,* vol. 230, no. 2, pp. 329-337, 2004.

[225] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE reviews in biomedical engineering,* vol. 10, pp. 213-234, 2017.

[226] D. Saranyaraj, M. Manikandan, and S. Maheswari, "A deep convolutional neural network for the early detection of breast carcinoma with respect to hyper-parameter tuning," *Multimedia Tools and Applications,* vol. 79, no. 15, pp. 11013-11038, 2020.