NEW CLASSIFIER ARCHITECTURE AND TRAINING METHODOLOGIES FOR LUNG NODULE DETECTION IN CHEST RADIOGRAPHS AND COMPUTED TOMOGRAPHY

Dissertation
Submitted to
The School of Engineering of the UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for
The Degree of
Doctor of Philosophy in Engineering

By
Barath Narayanan Narayanan
Dayton, Ohio
December, 2017
NEW CLASSIFIER ARCHITECTURE AND TRAINING METHODOLOGIES FOR
LUNG NODULE DETECTION IN CHEST RADIOGRAPHS AND COMPUTED
TOMOGRAPHY

Name: Narayanan, Barath Narayanan

APPROVED BY:

________________________________________
Russell C. Hardie, Ph.D.
Advisory Committee Chairman
Professor
Electrical and Computer Engineering

________________________________________
Eric J. Balster, Ph.D.
Committee Member
Associate Professor
Electrical and Computer Engineering

________________________________________
Vijayan K. Asari, Ph.D.
Committee Member
Assistant Professor
Electrical and Computer Engineering

________________________________________
Temesguen M. Kebede, Ph.D.
Committee Member
Research Professor
Electrical and Computer Engineering

________________________________________
Muhammad Usman, Ph.D.
Committee Member
Associate Professor
Mathematics

________________________________________
Robert J. Wilkens, Ph.D., P.E.
Associate Dean for Research and Innovation
Professor
School of Engineering

________________________________________
Eddy M. Rojas, Ph.D., M.A., P.E.
Dean, School of Engineering
© Copyright by

Barath Narayanan Narayanan

All rights reserved

2017
ABSTRACT

NEW CLASSIFIER ARCHITECTURE AND TRAINING METHODOLOGIES FOR LUNG NODULE DETECTION IN CHEST RADIOGRAPHS AND COMPUTED TOMOGRAPHY

Name: Narayanan, Barath Narayanan
University of Dayton

Advisor: Dr. Russell C. Hardie

Early detection of pulmonary lung nodules plays a significant role in the diagnosis of lung cancer. Radiologists use Computed Tomography (CT) and Chest Radiographs (CRs) to detect such nodules. In this research, we propose various pattern recognition algorithms to enhance the classification performance of the Computer Aided Detection (CAD) system for lung nodule detection in both modalities. We propose a novel optimized method of feature selection for clustering that would aid the performance of the classifier. We make use of an independent CR database for training purposes. Testing is implemented on a publicly available database created by the Standard Digital Image Database Project Team of the Scientific Committee of the Japanese Society of Radiological Technology (JRST). The JRST database comprises 154 CRs containing one radiologist confirmed nodule in each. We make use of 107 CT scans from publicly available dataset created by Lung Image Database Consortium (LIDC) for this study. We compare the performance of the cluster-
classifier architecture to a single aggregate classifier architecture. Overall, with a specificity of 3 false positives per case on an average, we show a classifier performance boost of 7.7% for CRs and 5.0% for CT scans when compared to single aggregate classifier architecture.

Furthermore, we study the performance of a CAD system in CT scans as a function of slice thickness. We believe this study has implication for how CT is acquired, processed and stored. We make use of CT cases acquired at a thickness of 1.25mm from the publicly available Lung Nodule Analysis 2016 (LUNA16) dataset for this research. We study the CAD performance at a native thickness of 1.25mm and various other down-sampled stages. Our study indicates that CAD performance at 2.5mm is comparable to 1.25mm and is much better than at higher thicknesses. In addition, we propose and compare three different training methodologies for utilizing non-homogenous thickness training (i.e., composed of cases with different slice thicknesses). We utilize cases acquired at 1.25mm and 2.5mm respectively from the LUNA16 dataset for this study. These methods include: (1) aggregate training using the entire suite of data at their native thickness, (2) homogeneous subset training that uses the subset of training data that matches each testing case; and (3) resampling all training and testing cases to a common thickness. Our experimental results indicate that resampling all training and testing cases to 2.5mm provides the best performance among the three training methods compared. Furthermore, the resampled 2.5mm data require less memory and process faster than the 1.25mm data.
Dedicated to Appa and Amma
ACKNOWLEDGEMENTS

I take this opportunity to thank Dr. Russell C. Hardie for serving as my Ph.D. advisor and mentor. His support and guidance in research has truly been remarkable. He instilled confidence in me at key moments. He has motivated me throughout my journey as a graduate student and will continue to do so in the years to come. I extend my thanks to Dr. Guru Subramanyam for supporting me financially throughout my Ph.D. journey. I thank Dr. Temesguen M. Kebede for being a great support and providing wonderful insights. I also express my gratitude to other committee members, Dr. Vijayan Asari, Dr. Eric Balster and Dr. Muhammad Usman for their helpful discussions. Additionally, I thank Graduate Academic Affairs Office of University of Dayton for assisting me through Graduate Student Summer Fellowship program in 2015 & 2017. I also thank Matthew Sprague for his valuable inputs during our research meetings.

I render my sincere thanks to the publicly available Computed Tomography databases created by Lung Imaging Database Initiative & Image Database Resource Initiative along with Lung Nodule Analysis (LUNA16). Thanks to scientific committee of Japanese Society of Radiological Technology for creating publicly available chest radiography dataset. These databases helped me complete my research and evaluate my performance among my peers across the world.

Thanks to a very special person in my life, Priyamvada, for being a pillar of support since the start of my Ph.D. journey. Finally, I would like to attribute all my success to my
parents, Narayanan and Sukanya. They made me what I am today and without them, I would have not been able to accomplish anything. The faith, love, support and sacrifices that both my mom and girlfriend have gone through for the completion of my doctorate degree has been immeasurable and they deserve every bit of this success.
TABLE OF CONTENTS

ABSTRACT ........................................................................................................... iv

DEDICATION ...................................................................................................... vi

ACKNOWLEDGEMENTS ................................................................................ vii

LIST OF FIGURES ........................................................................................ xi

LIST OF TABLES ............................................................................................ xv

CHAPTERS

1. INTRODUCTION ...................................................................................... 1

2. MATERIALS ................................................................................................ 7
   2.1. Chest Radiograph Materials ............................................................... 7
   2.2. Computed Tomography Materials .................................................... 9
       2.2.1. LIDC-IDRI Dataset ................................................................ 9
       2.2.2. LUNA16 Dataset ................................................................ 11

3. CAD SYSTEM FRAMEWORK .................................................................. 12
   3.1. Potential Candidate Detector ........................................................... 12
       3.1.1. Candidate Detector for CRs ..................................................... 12
       3.1.2. Candidate Detector for CT Scans .......................................... 17
   3.2. CAD Classification System ............................................................. 18
       3.2.1. Feature Computation ............................................................. 19
LIST OF FIGURES

2.1. Chest radiograph example from the JRST dataset............................... 7
2.2. Lung nodule marked by one of the radiologists for a chest radiograph from

JRST dataset ........................................................................................................ 8
2.3. CT scan example from the LIDC-IDRI dataset............................................. 10
2.4. Lung nodule marked by one of the radiologists for a CT scan from the

LIDC–IDRI dataset ................................................................................................. 10
3.1. Block diagram of the potential candidate detector system for CR..................12
3.2. LCE output for a typical chest radiograph from the JRST dataset.................... 14
3.3. Lung Segmentation result for a chest radiograph from the JRST dataset......... 15
3.4. Typical candidate detector result for a chest radiograph from the JRST dataset ...... 16
3.5. ADT Segmentation result for a potential nodule candidate ............................. 17
3.6. Block diagram of the potential candidate detector system for CT scans............ 17
3.7. Typical candidate detector result for a CT scan from the LIDC-IDRI dataset........ 18
3.8. Block diagram of the CAD classification system for chest radiographs

and CT scans........................................................................................................ 18
4.1. Block diagram of the aggregate classification method.................................. 22
4.2. Block diagram of the intuitive cluster classification method ......................... 24

4.3. Block diagram of the optimized cluster classification method ....................... 25

4.4. Block diagram representing both intuitive and optimized cluster method of classification ................................................................. 27

4.5. Potential candidate detector result for a chest radiograph testing case ............. 28

4.6. Selection of classifier features for aggregate classification method using
Riverain dataset .................................................................................. 29

4.7. Nodules clustered differently using intuitive cluster features .......................... 33

4.8. Selection of classifier features for intuitive cluster classification method using
Riverain dataset .................................................................................. 34

4.9. Selection of optimized cluster features using Riverain dataset ....................... 35

4.10. Nodules clustered differently using optimized cluster features for classification ... 38

4.11. Selection of classifier features for optimized cluster classification method
using Riverain dataset ......................................................................... 39

4.12. FROC comparison of the classification performance for chest radiographs
utilizing different classification architectures ....................................... 40

4.13. CAD System output for the case ‘JPCLN018’ using aggregate and
optimized cluster classification method .............................................. 41
4.14. CAD System output for the case ‘JPCLN031’ using aggregate and optimized cluster classification method .............................................. 42

4.15. FROC comparison of the CAD classification performance for CT scans utilizing different classification architectures .......................................................... 44

5.1. Typical CT scan viewed at different slice thicknesses from the LUNA16 dataset .... 48

5.2. Image of small nodule at different thicknesses .......................................................... 51

5.3. Image of large nodule at different thicknesses .......................................................... 53

5.4. Selection of classifier features at all thickness stages .............................................. 54

5.5. FROC comparing CAD performance at different thickness levels ...................... 54

5.6. CAD system output for the case ‘sub0_p32’ at different thicknesses .................. 57

5.7. Histogram distribution of size and contrast of the nodules successfully detected by CAD system at different thicknesses .............................................. 58

6.1. Selection of classifier features using different training methods for

1.25mm testing dataset .................................................................................................. 61

6.2. FROC comparing overall CAD performance using different training methods for 1.25mm testing dataset .............................................................................. 63

6.3. Selection of classifier features using different training methods for

2.5mm testing dataset .................................................................................................. 65
6.4. FROC comparing overall CAD performance using different training methods for 2.5mm testing dataset ................................................................. 67

6.5. FROC based on 10-fold validation using different training methods for LUNA16 dataset ................................................................. 68
LIST OF TABLES

2.1. Distribution of LUNA16 dataset utilized for this research.......................... 11

4.1. Classification performance of the CAD systems at 3 FPs for CR using the
three classification architectures................................................................. 43

4.2. Classification performance of the CAD systems at 3 FPs for CT modality
using the three classification architectures.................................................. 45

5.1. Overall CAD performance comparison at all thickness levels..................... 55

6.1. Training and testing dataset compositions for different methods of
classification for 1.25mm testing dataset.................................................. 60

6.2. Overall CAD performance comparison using different training methods
for 1.25mm testing dataset............................................................................. 62

6.3. Training and testing dataset compositions for different methods of
classification for 2.5mm testing dataset..................................................... 64

6.4. Overall CAD performance comparison using different training methods
for 2.5mm testing dataset............................................................................. 66
According to the National Cancer Institute, 158,080 lung cancer cases were reported by the end of 2016 [1]. World Health Organization (WHO) report in the year 2014 states that lung cancer is the most common cause of cancer-related deaths [2]. Studies have proven that early detection of lung cancer increases one’s survival rate (87% if diagnosed in Stage I) [3]. Lung cancer usually exhibits its presence with the formation of pulmonary nodules. Early detection of such cancerous nodules could improve one’s chance of survival [4]. Detection of such potentially cancerous nodules is a problem attracting great interest. These nodules are ellipsoidal lesions present in the lung with well-defined boundaries. Imaging techniques such as Computed Tomography (CT) and Chest Radiographs (CRs) are used to detect such cancerous nodules. In the Early Lung Cancer Action Project, low dose CT has proven to be more effective than CRs [4]. These imaging techniques, especially low-dose CT, have been successful in the initial screening of lung cancer detection. In high resolution CTs, it is not only possible to identify such nodules but also make some useful measurements regarding its shape, volume, etc. CT provides numerous slices of image data which can be time consuming and potentially fatiguing for radiologists to study. Also, dealing with this bulk volume of CT data is quite a challenge. On the other hand, CRs are widely prescribed, simple and relatively less expensive. In addition, CRs
provide only a fraction of X-ray dose when compared to CT scans. Studies have shown that 90\% of peripheral lung cancer nodules are visible in CRs [5]. Furthermore, Food and Drug Administration (FDA): RS-2000 (Deus Technologies, Rockville, MD) has approved a Computer Aided Detection (CAD) system for detection of pulmonary nodules in CRs. FDA approval suggests that automated analysis of such data is very essential. Hence, CAD system to detect such cancerous nodules in both these modalities would provide a valuable second opinion and enhance the work flow of a radiologist.

CAD of lung nodules in CRs and CT scans has been a research area attracting great interest for the last few decades. Several CAD research papers have been presented in the literature [6-43]. For instance, [6] proposed a novel CAD system for identifying lung nodules in CRs that comprised of a novel Weighted Multiscale Convergence Index (WMCI) filter that detects all the potential nodule candidates and an Adaptive Distance based Threshold (ADT) algorithm for its segmentation. A set of 114 features is computed for each candidate. Gaussian Bayes linear classifier, Fisher Linear Discriminant (FLD) and quadratic classifier are compared in [6] as well. In [7], a novel CAD system to detect lung nodules in CT scans is presented. Detection and segmentation of the potential nodule candidates is done simultaneously using intensity thresholding with morphological processing. A set of 245 features were computed for every potential nodule candidate and the best results for classification were obtained using FLD classifier. ‘N-Quoit filter’ is utilized in [8]. The diagnosis rules based on fuzzy clustering is described in [9]. In 1999, [10], a hybrid classifier combining an unsupervised and a supervised model was designed to improve classification performance of malignant and benign masses in Mammograms. In 2001, an improved system that combines 2-D and 3-D feature analysis along with linear
discriminant classifier was proposed in [11]. Another system [12] was designed based on improved template-matching technique built on genetic algorithm. New gradient concentration features for recognizing polyps was proposed in [13]. A detection system that uses a surface normal overlap technique along with simple rule-based classifier to attenuate False Positive (FP) findings is presented in [14]. Various classification techniques such as support vector machine, kernel Fisher discriminant and AdaBoost for automated classification were applied in [15]. A localized search method based on anatomical classification to detect potential nodule candidates in CRs is used in [16]. A dot enhancement filter for candidate selection and a neural classifier for reduction of FPs is incorporated in [17]. In [18], K-means clustering is applied for the initial detection and segmentation of nodules and later divided into six groups based on its thickness and percentage of connectivity with lung walls. Certain fixed set of 2-D and 3-D features are extracted from each candidate based on its type for classification purposes. In [19], Automated Nodule Detection (ANODE09) database from lung cancer screening and a web-based program for evaluation of nodule detection algorithms is presented. A method to combine the results of various CAD system results is also proposed in [19]. A novel rolling ball algorithm for lung segmentation to improve the detection of juxtapleural nodules is presented in. Multi-level thresholding is implemented for nodule detection. The National Cancer Institute launched a new database in [21] to simulate the advancement of CAD of lung nodules. National Cancer Institute cooperated with five other academic institutions to create Lung Image Database Consortium – Image Database Resource Initiative (LIDC-IDRI). We make use of this publicly available dataset for our research as well. Chapter 2 provides further description about the LIDC-IDRI dataset utilized in this research. CAD
performance in CT scans at standard dose and ultra-low dose is compared in [22]. In [23],
two CAD systems on detection of small pulmonary nodules at multi-detector row CT scans
are compared using a consensus panel as reference. Evaluation of CAD performance using
low-dose CT scan is implemented in [24]. In [25], automated detection of lung nodules is
implemented for the same set of patients using different slice thickness protocols in
multidetector CT. In, [26] a CAD system is presented for LIDC-IDRI dataset using the
Free Receiving Operating Curve (FROC) analysis thereby setting a benchmark. In [27], an
experiment is performed to compare the detection of nodules by CAD system depending
on its size. Evaluation of CAD systems using artificial intelligence techniques for the
detection of pulmonary embolism is presented in [28]. Finally in [29], potential nodule
candidates are determined by using multi-scale techniques which includes local intensity
maxima in Gaussian scale space. In [29], classification performance of the CAD system is
improved by reduction of potential nodule candidates and better candidate segmentation.
Some of the other published algorithms are described in [30-43].

Nevertheless, there is still scope for further improvement. A typical CAD system
contains a potential candidate detector, candidate segmentator, and a feature based
classifier to designate candidates as nodules or non-nodules. One area for improvement,
we believe, is within the design of feature based classifiers. A typical CAD feature set tends
to have numerous diverse candidates thereby creating statistical differences among the
candidates belonging to the same category. Hence, we believe that bifurcation of data by
means of clustering would improve the overall performance of the CAD system. Clustering
has improved the classification performance for various applications surveyed in [30].
Except for [18], cluster based classification has not been widely studied for CAD systems.
We reckon this is an important area to investigate. In this research, we present a novel method of optimized feature selection for both clustering and classification. A CAD algorithm is usually presented for a specific modality but in this research, we demonstrate the efficacy of the proposed algorithm for both CRs and CT scans. Performance of our CAD system in CRs is measured on a publicly available dataset created by the Standard Digital Image Database Project Team of the Scientific Committee of the Japanese Society of Radiological Technology (JRST) [31]. For CT scans, the cluster based classifier performance is studied on the publicly available LIDC dataset.

In addition to cluster-based architecture, we address two important issues for CAD of lung nodules in CT scans. The first issue relates to how slice thickness impacts CAD performance given training and testing data are of same thickness. This experiment has implications for how CT is acquired and/or how it may be resampled for CAD processing. The second issue relates to how to train a CAD system for best performance given non-homogenous slice thickness training data. Generally, one would like to use all the training data available. However, this would mean pooling of CT scans obtained from variety of scanners and acquisition parameters, such as slice thickness and dosage settings. We propose and compare three methodologies for utilizing non-homogenous slice thickness training data.

To study the impact of slice thickness on CAD performance, we use the following approach. We study the CAD performance at native thickness of 1.25mm and three other down-sampled stages for the same set of training and testing cases. This study would help us determine the slice thickness at which a CT scan could be acquired for optimal CAD performance both in terms of accuracy and computational complexity. To determine the
best method of training for non-homogenous slice thickness data, we propose and compare three methodologies. At first, we employ the traditional CAD system approach where the entire suite of data is utilized at their native thickness (aggregate training method). Later, we study a homogenous approach where only the cases that match with the slice thickness of testing data would be utilized for training purposes. Finally, we re-sample all the training and testing cases to a specific thickness value and study its impact on CAD performance as well. All the experiments conducted in this research are implemented on the publicly available Lung Nodule Analysis (LUNA16) dataset [32, 33] thereby setting a benchmark for future research efforts. We conduct two sets of experiments utilizing 1.25mm and 2.5mm slice thickness data from LUNA16 dataset for each of the three methodologies to validate our study.

The remainder of the dissertation is organized as follows. Chapter 2 provides a brief description about the databases that are employed for this research. Chapter 3 describes the CAD system architecture adopted in this research. This includes the description of the potential candidate detector system utilized along with the features extracted for both CRs and CT scans. Chapter 4 presents the cluster based classification architecture. Chapter 5 describes the study of CAD system as a function of slice thicknesses. Chapter 6 elucidates the various training methods with non-homogenous data. Finally, conclusions are offered in Chapter 7, in addition to potential avenues for future research.
2.1. Chest Radiograph Materials

For CRs, we make use of separate databases to study the performance of our CAD system as implemented in [6]. Training of the CAD system for CRs is implemented on the independent dataset provided by Riverain Medical [6]. The training dataset comprises 160 CRs with 173 lung nodules [6].

Figure 2.1: Chest radiograph example from the JRST dataset
To measure the performance of our CAD system in CRs, we make use of publicly available JRST dataset. This dataset comprises 154 chest radiographs with lung nodule detection reviewed by a panel of three radiologists and consequently verified in CT data. The scans are of size $2048 \times 2048$. Pixel spacing of 0.175mm and 4096 intensity levels. Every case in JRST comprises one radiologist confirmed nodule in each. Among the 154 nodules present in JRST dataset, 100 are identified as malignant and 54 as benign. In addition to nodule center coordinates, other information such as nodule size, malignancy rating, subtlety rating and patient information are provided. Figure 2.1 shows a typical CR (Case: JPCLN001) provided in the JRST dataset along with a nodule marked by one of the panel radiologists. Figure 2.2 shows the lung nodule marked by one of the radiologists for the case JPCLN001 from the JRST dataset. Utilizing independent datasets for training and testing is a testimony to robustness of our CAD system.
2.2. Computed Tomography Materials

2.2.1. LIDC – IDRI Dataset

LIDC-IDRI is a publicly available dataset in The Cancer Imaging Archive (TCIA) created for the development of CAD systems in CT scans. The LIDC-IDRI data is collected from various sites within the United States. This established database was initiated by the National Cancer Institute (NCI), which was further enhanced by the foundation of the National Institutes of Health (FNIH) along with FDA. The LIDC-IDRI dataset contains 1018 CT scans of 1010 different patients and each of them are studied by at least one of the four radiologists. In the initial blind-read phase, each radiologist annotated the CT scan independently and marked suspicious lesions. Later, each radiologist reviewed their own marking independently with the other three radiologists’ markings to render their final opinion. There is some dissent among the radiologists about many of their nodule findings even after the second session of study. This database contains 7371 lesions marked as nodule by at least one of the radiologists out of which 2669 are above 3mm. The detection results include the markings made by each radiologist, its size, segmentations and its corresponding rating. Figure 2.3 shows a slice of a case (LIDC-IDRI: 0001) from the LIDC-IDRI dataset along with the radiologists’ markings. Figure 2.4 shows a nodule marking by one of the radiologists for the case LIDC-IDRI 0001.
Figure 2.3: CT scan example from the LIDC-IDRI dataset

Figure 2.4: Lung nodule marked by one of the radiologists for a CT scan from the LIDC

– IDRI dataset
2.2.2. LUNA16 Dataset

LUNA16 is a grand challenge set up for the evaluation of CAD algorithms for detection of lung nodules [32, 33]. Dataset used for this challenge is a subset of LIDC-IDRI database provided at the National Biomedical Imaging Archive. For the LUNA16 grand challenge, 888 CT scans are selected from the LIDC-IDRI dataset. A panel of four radiologists studied CT scans present in the LUNA16 dataset. These 888 CT scans comprises 1351 radiologists’ markings. Till date, not much research has been done on this dataset. The LUNA16 dataset contains substantial quantity of CT scans with different slice thicknesses, which is ideal for the thickness study conducted in this research. Table 2.1 enlists the slice thickness distribution of the LUNA16 dataset utilized for this research. We utilize 192 cases from the LUNA16 dataset with 268 nodule cue points (no redundancy) with slice thickness and spacing of 1.25mm. In addition, we make use of 283 cases with 322 nodule cue points (no redundancy) marked by radiologists at 2.5mm slice thickness and spacing.

Table 2.1: Distribution of LUNA16 dataset utilized for this research

<table>
<thead>
<tr>
<th>Slice Thickness &amp; Spacing (mm)</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25</td>
<td>192</td>
</tr>
<tr>
<td>2.5</td>
<td>283</td>
</tr>
</tbody>
</table>
Automated CAD of lung nodules has been an active field of research and several CAD systems have been proposed in the literature [4-43]. In this chapter, we describe the CAD system architecture adopted in this research for CR and CT modalities respectively. At first, we briefly describe the potential candidate detector of the CAD system for both CR and CT scans before we shift our emphasis to the classification process.

3.1. Potential Candidate Detector

3.1.1. Candidate Detector for CRs

Figure 3.1 shows the block diagram of the potential candidate detector system for CRs adopted from [6].

![Block diagram of the potential candidate detector system for CR](image)

**Figure 3.1**: Block diagram of the potential candidate detector system for CR [6]
For this research, we focus on candidate detector systems proposed in [6] which is performed using the following methods (i) image re-sampling (ii) Local Contrast Enhancement (LCE) (iii) lung segmentation (iv) detection of potential nodule candidates and (v) candidate segmentation. In image re-sampling step, all the input images are re-sampled to a common pixel spacing of 0.7 mm. Re-sampling of images helps in noise smoothing and significant reduction of computational time. This also helps in maintaining the resolution compatibility between the training and testing data.

The operation of LCE is performed to normalize the contrast across different images and within each image. This process is given by:

\[
y(m, n) = \frac{x(m,n) - \mu(m,n)}{\sigma(m,n)},
\]

(1)

where \(x(m, n)\) represents the down-sampled image, \(y(m, n)\) shows the LCE output image, \(\sigma(m, n)\) is the local standard deviation estimate and \(\mu(m, n)\) is the local mean estimate which is computed using a Gaussian low pass filter. Equation 2 shows the formula used for the computation of local mean estimate and is given by

\[
\mu(m, n) = x(m, n) * h(m, n),
\]

(2)

where * represents 2-D convolution. \(\sigma(m, n)\) is calculated as follows:

\[
\sigma(m, n) = (x^2(m, n) * h(m, n) - \mu^2(m, n))^{0.5},
\]

(3)

where kernel \(h(m, n)\) impacts the type and scale of the objects enhanced using LCE with a standard deviation of 16. Figures 3.2(a) and 3.2(b) illustrate the difference between the images before and after the application of LCE.
Figure 3.2: LCE output for a typical chest radiograph (Case: JPCLN019) from the JRST dataset (a) Raw image and (b) LCE image
Lung segmentation is executed on the down-sampled image as described in [6] using an Active Shape Model (ASM). Figure 3.3 illustrates the lung segmentation for the case JPCLN019 from the JRST database after the application of LCE. Any nodule present outside the lung field is missed by our CAD system.

![Lung Segmentation Result](image)

**Figure 3.3**: Lung Segmentation result for a chest radiograph from the JRST dataset (Case: JPCLN019)

Potential nodule candidates are determined using a WMCI filter [35, 36]. The parametric values for WMCI filter are chosen as mentioned in [6]. Basic idea of CI filter is that structures in images with circular shape that are brighter than their surroundings will tend to have gradients that point towards the center of the object. Figure 3.4 shows a typical potential candidate detector result obtained after the application of WMCI filter.
Candidate segmentation is implemented using ADT algorithm as described in [6]. This algorithm uses the LCE image and thresholding is done by using Equations 4-6. \( s(m, n, T_0) \) represents the binary image. \( T(m, n, T_0) \) is the adaptive threshold function of the distance \( d \) of a given pixel \( (m, n) \) from the detection cue point \( (m_0, n_0) \). \( T_0 \) is the adaptive threshold parameter. \( T_\Delta \) determines the range of thresholds from the cue point to those at maximum radius \( r_{max} \). Figure 3.5 shows a candidate segmented using ADT segmentation algorithm.

\[
\begin{align*}
    s(m, n, T_0) &= \begin{cases} 
    1 & y(m, n) > T(m, n, T_0) \\
    0 & y(m, n) \leq T(m, n, T_0) 
\end{cases} \\
    T(m, n, T_0) &= \begin{cases} 
    T_0 + T_\Delta \left( 1 - \frac{d(m, n)}{r_{max}} \right) & d(m, n) < r_{max} \\
    0 & otherwise
\end{cases}
\end{align*}
\]

\[
d(m, n) = (m - m_0)^2 + (n - n_0)^2
\]
3.1.2. Candidate Detector for CT Scans

The first step in preprocessing stage is to orient the CT scans consistently. Figure 3.6 shows the block diagram of the potential candidate detector for CT scans. 3-D lung segmentation is performed on the oriented CT scans as described in [7]. Later, potential nodule candidates are detected and segmented simultaneously using intensity based
thresholding with morphological processing [7]. Figure 3.7 shows a typical candidate detector result.

![Figure 3.7: Typical candidate detector result for a CT scan from the LIDC-IDRI dataset (Case: LIDC-IDRI-0001)](image)

**3.2. CAD Classification System**

![Figure 3.8: Block diagram of the CAD classification system for CR and CT scans from [6] and [7]](image)
3.2.1. Feature Computation

After the identification of potential nodule candidates, the CAD system needs to perform a pattern recognition task. Figure 3.8 shows the block diagram of the CAD classification system. To implement a pattern recognition task, the candidates are represented by points in feature space. Various set of features have been proposed to enable the classifier to distinguish between nodules and non-nodules. Most of these set of features have been derived based on histogram information [37, 38], candidate shape [39, 40] and filter outputs [41]. In this research, we employ the feature extraction techniques described in [6] and [7] for CR and CT scans respectively.

In [6], for each potential nodule candidate, a set of 114 features are computed. These include a set of 9 geometric features, 18 intensity features and 17 gradient features. Geometric features are computed based on the information provided by the ADT algorithm. Intensity and gradient features are extracted for normalized, LCE and WMCI images.

For CT scans, we compute an entire suite of 503 features for all the potential candidates detected which includes the 345 features mentioned in [7]. The 345 features are shortlisted to 245 based on linear independence in [6]. We differ in this research by shortlisting the top 300 based on rank with Receiver Operating Characteristics (ROC) curve criterion [44].

3.2.2. Feature Selection

Later, a subset of features is used by the classifier to distinguish the candidates. The subset of features is determined using Sequential Forward Selection (SFS) method proposed in [45]. In SFS, features are added to an empty set one-by-one. At each step, one
feature is added and we measure the performance of the system. The feature that performs the best is selected and the process is repeated. The main aim of this step is to select features that would maximize the performance for a given classifier. This type of selection would help us avoid exhaustive enumeration.

3.2.3. Performance Measure

ROC is a 2-D depiction of accuracy of a detector for a given set of testing data. It plots the probability of detection of a nodule (i.e. True Positive (TP) rate) versus the probability of false alarm (i.e. FP rate). Each point in the curve corresponds to different decision threshold. As one gradually moves towards the right of the curve, the decision metric changes from strict to lenient.

FROC is a modified version of the ROC. FROC measures the overall sensitivity of the CAD system for a set of average number of FPs per cases. Overall sensitivity of the CAD system is the measure of accuracy of detection of the entire CAD system (including the potential candidate detection and classification). In this research, we estimate the performance by measuring the area under the FROC between 0-10 FPs for both CRs and CT scans.

3.2.4. FLD Classifier

In this research, we primarily focus on FLD classifier. A FLD classifier forms a linear decision boundary to distinguish the patterns. It is a special case of linear classifier [45]. Here, we have a common covariance for all the samples instead of individual covariance for respective classes. Common covariance is estimated based on the mean of covariance of the respective classes as shown:

$$\Sigma = \frac{1}{2} (\Sigma_1 + \Sigma_2).$$  \hspace{1cm} (7)
The decision of a FLD classifier is based on a discriminant function \( g \) which is shown below:

\[
g(x) = (M_2 - M_1)^T \Sigma^{-1} x + 0.5 \times (M_1^T \Sigma^{-1} M_1 - M_2^T \Sigma^{-1} M_2) + \log \left( \frac{p_2}{p_1} \right). \tag{8}
\]

Here, \( M_1 \) and \( M_2 \) represent the mean of their respective classes. \( x \) represents the test data in question. \( P_1 \) and \( P_2 \) represent the prior probability of both the classes.

Corresponding class labels are given as follows:

\[
\text{Class Assignment} = \begin{cases} 
1 & g(x) \leq 0 \\
2 & g(x) > 0 
\end{cases}
\tag{9}
\]

We embark upon different approaches in Chapters 4 to 6 that enhance the performance of the classification system and thus eventually improving the nodule detection rate.
CHAPTER 4

FEATURE SELECTION BASED CLUSTERING APPROACH

In this chapter, we describe the various classification architectures implemented in this research for comparing the performance of an aggregate with cluster-based classification approach.

4.1. Aggregate Classification

Figure 4.1: Block diagram of the aggregate classification method

Figure 4.1 shows the block diagram of the aggregate classification method implemented in both [6] and [7]. In CR, train and test data refers to 114 feature values of all the potential candidates detected by the candidate detector for Riverain and JRST datasets respectively. However, for CT scans, we make use of 300 feature values.

Let $N$ and $M$ represent the total number of features and total number of candidates detected by the potential candidate detector for the training dataset, respectively.
A feature matrix that contains all the feature values for every potential candidate detected by the candidate detector for the training dataset is represented by $\mathbf{X}$ and is of size $N \times M$. In both [6] and [7], SFS of features is implemented to determine optimal set of classifier features. Let $P$ represent the number of features selected by SFS and $P \leq N$. A specific set of feature indices $\emptyset$ is defined by SFS, of size $P \times 1$, given by

$$\emptyset = SFS_A(\mathbf{X}).$$ (10)

Operator $SFS_A$ represents a SFS function for aggregate classification technique where $SFS_A: \mathbb{R}^{N \times M} \rightarrow Q^P$. The classification sensitivity of the CAD system for a set of average number of FPs per case is measured in terms of FROC. A point to note, for all the SFS methods mentioned in this research (CR and CT scans), performance is determined based on AUC between 0-10 FPs obtained during 10-fold cross-validation of the training dataset considered. The train feature matrix $\mathbf{X}$ is translated to a set of feature indices $\emptyset$ using SFS and is defined as

$$\emptyset = \{\emptyset_1, \emptyset_2, \ldots, \emptyset_P\},$$ (11)

where $\emptyset \subseteq \{1, 2, \ldots, N\}$. Let $\mathbf{X}(\emptyset)$ represent the matrix of features from the set $\emptyset$ for all candidates. Final detections of the CAD system are provided using a classifier based only on $\mathbf{X}(\emptyset)$. The candidates are designated as nodules or non-nodules with the help of a FLD classifier. It has the capability to form a well-defined boundary even with uneven distribution of data. Aggregate classifier results are presented in Chapter 4.4.
4.2. Intuitive Cluster Classification

Figure 4.2 presents the block diagram of the intuitive cluster classification method discussed in this research. In this method, we pick certain features based on our intuition for clustering purposes and the feature set is represented by $I$. K-means method of clustering is opted in this research because it has proven to be highly effective, fast and computationally simple. K-means clustering provides the cluster labels represented by $L_I$, for all candidates detected by the candidate detector. Cluster labels ($L_I$) are of size $M \times 1$ and is given by

\[ L_I = Kmeans(X(I), K). \] (12)

Here, $Kmeans$ represents the K-means clustering process using only the intuitive feature values of every candidate ($X(I)$) and $K$ represents the number of clusters. Clusters
for the testing data are assigned based on the smallest Euclidean distance from the cluster centers. After clustering, we optimize the classifier features for each cluster by performing SFS on the clustered data individually and is given by

$$\left[\emptyset_{I,1}, \emptyset_{I,2}, ... \emptyset_{I,K}\right] = SFS_{C}(X, L_I).$$

(13)

Operator $SFS_{C}$ represents the feature selection process for the clustered data determined based on $L_I$. Output of the Operator $SFS_{C}$, $\emptyset_{i,j}$, represents the set of classifier feature indices selected for cluster $j$ after clustering based on intuitive features ($I$) where $j = 1, 2, ..., K$. Here, $X(\emptyset_{i,j})$ represents the matrix of features from the set $\emptyset_{i,j}$ for all candidates belonging to the $j^{th}$ cluster (clustered based on $I$). Later, we classify cluster $j$ based only on the $X(\emptyset_{i,j})$ data determined for that particular cluster.

4.3. Optimized Cluster Classification

![Block diagram of the optimized cluster classification method](image)

**Figure 4.3:** Block diagram of the optimized cluster classification method
Figure 4.3 shows the block diagram of optimized cluster classification method proposed in this research. For this method, we make use of the aggregate classifier features (Ø) determined as mentioned in Chapter 4.1. We implement SFS method of feature selection for clustering features ($SFS_K$) by fixing Ø for classification i.e., K-means clustering of each feature is implemented individually and 10-fold classification performance using the aggregate classifier features (Ø) is studied. Cluster features that provide the best results are noted and this process is repeated until a good suite of features is determined. The process of selecting features for clustering is given by

$$\varphi = SFS_K(X, \emptyset, K).$$  \hspace{1cm} (14)

Operator $SFS_K$ represents the process of SFS for clustering features and $\varphi$ represents the optimal cluster feature indices determined. This type of cluster feature selection would make sure that overall classifier performance does not drop due to clustering as we monitor its performance using aggregate classifier features. Feature matrix ($X$) is later clustered using $\varphi$ and is given by

$$L_\varphi = Kmeans(X(\varphi), K).$$  \hspace{1cm} (15)

Here, $L_\varphi$ represents the cluster labels provided by K-means clustering process using $X(\varphi)$ and $K$, where $X(\varphi)$ are optimal cluster feature values of every candidate. Later, we repeat the classification process like the one described in Chapter 4.2. Test data cluster is determined based on the closest cluster center. The sets of classifier features for each cluster are given by

$$[\emptyset_{\varphi,1}, \emptyset_{\varphi,2}, \ldots, \emptyset_{\varphi,K}] = SFS_C(X, L_\varphi).$$  \hspace{1cm} (16)
The set \( \emptyset_{\varphi,j} \) represents the classifier feature indices selected for cluster \( j \) after clustering based on optimal cluster features (\( \varphi \)). Here \( X(\emptyset_{\varphi,j}) \) represents the matrix of features from the set \( \emptyset_{\varphi,j} \) for all candidates belonging to cluster \( j \) (clustered based on \( \varphi \)). We classify each cluster individually using \( X(\emptyset_{\varphi,j}) \) determined for cluster \( j \) with the help of a FLD classifier.

Figure 4.4 illustrates the difference between the intuitive cluster classification and optimized method of cluster classification. At the current position, it depicts the intuitive cluster classification and when the switch position is altered, it resembles optimized cluster classification method.

Figure 4.4: Block diagram representing both intuitive and optimized cluster method of classification
4.4. Experimental Results for Feature Selection Based Clustering Approach

4.4.1. Chest Radiograph Results

Figure 4.5: Potential candidate detector result for a chest radiograph testing case (Case: JPCLN028)

Figure 4.5 illustrates the results for the case ‘JPCLN028’ from the JRST database after the application of potential candidate detector described in Chapter 3. Candidate detector sensitivity is found to be 93.1% and 95.0% for Riverain and JRST respectively, i.e., it was successfully able to detect 161 and 133 nodules among the 173 and 140 from each dataset as obtained in [6].
Figure 4.6: Selection of classifier features for aggregate classification method using Riverain dataset

Figure 4.6 presents the SFS merit function based on 10-fold validation of the Riverain dataset for aggregate classification method. The selection of operating point from the AUC plot for the number of features for classification purposes is implemented as in [6, 7]. In this scenario, 26 features are selected and classification is performed exclusively based on those features. We present the classification results later in this chapter.

Based on empirical study, we determined that two clusters are optimal for cluster classification purposes. In addition, 2-cluster division makes sure there is sufficient distribution of nodules in both the clusters to provide consensus among statistics. For intuitive cluster classification, we pick Size, Eccentricity and Maximum Raw Intensity candidate features for clustering. Size contributes to the geometrical aspect, eccentricity
for circular nature and intensity tells how bright the candidate is. Figure 4.7 presents six nodules clustered differently based on intuitive features from the JRST dataset. In Figure 4.7, (a), (b) and (c) belong to a cluster category whereas the remaining are associated with the other cluster. In Figure 4.7, the segmented nodules (a), (b) and (c) clearly appear to be brighter and larger in comparison to the ones in (d), (e) and (f) that exhibit circularity in addition. Figure 4.8 shows the SFS merit function for both the clusters. 17 and 16 features are selected for each cluster respectively.
Figure 4.7: Nodules belonging to different clusters based on intuitive features from JRST CR dataset: (a) JPCLN003 (Cluster I), (b) JPCLN005 (Cluster I), (c) JPCLN017 (Cluster I), (d) JPCLN019 (Cluster II), (e) JPCLN020 (Cluster II), and (f) JPCLN021 (Cluster II)
Figure 4.8: Selection of classifier features for intuitive cluster classification method using Riverain dataset

Figure 4.9 presents SFS merit function results obtained for clustering features based on 10-fold validation performance using 26 aggregate features for optimized cluster classification method. We pick an operating point of three features in this scenario. Figure 4.10 shows the six nodules clustered based on the selected features. In Figure 4.10, (a), (b) and (c) belong to a cluster category and they appear to be slightly more obvious when compared to the rest belonging to different cluster. However, visual differences between these nodules are not as prominent as observed in Figure 4.10 using intuitive cluster features. Figure 4.11 shows the SFS merit function obtained for classification features for both the clusters. Here too, we select 17 and 16 features for each cluster for classification purposes.
Figure 4.9: Selection of optimized cluster features using Riverain dataset
**Figure 4.10:** Nodules belonging to different clusters based on the cluster features selected using optimized cluster classification method from JRST CR dataset: (a) JPCLN015 (Cluster I), (b) JPCLN028 (Cluster I), (c) JPCLN029 (Cluster I), (d) JPCLN087 (Cluster II), (e) JPCLN026 (Cluster II), and (f) JPCLN031 (Cluster II)
Figure 4.11: Selection of classifier features for optimized cluster classification method using Riverain dataset
Overall FROC analysis for all the classification architectures is shown in Figure 4.12. As mentioned in Chapter 4, Riverain database is used for training purposes and JRST database is used for testing. Figure 4.12 clearly indicates that the optimized method of cluster classification performs the best among the classification methods studied in this research. Note that in [6], simple candidate cue adjacency rule is employed to avoid redundant candidates i.e., among the candidate detections within 22 mm from one another, only the candidate with the largest detection statistic is provided as the final system detection. We do not employ this rule for our results, which correlates to the lower performance when compared to [6]. However, our aggregate classification method matches with the performance mentioned in [6] without cue adjacency rule.
Figure 4.13: CAD system output for the case ‘JPCLN028’ using (a) aggregate classification method and (b) optimized cluster classification method
Figure 4.14: CAD system output for the case ‘JPCLN031’ using (a) aggregate classification method and (b) optimized cluster classification method
Figures 4.13 and 4.14 present typical CAD system output for the cases ‘JPCLN028’ and ‘JPCLN031’ provided by both (a) aggregate and (b) optimized cluster classification methods at an operating point of 3 FPs from the overall FROC curve. Figures 4.13 and 4.14 clearly indicate that the nodules that were missed by the traditional aggregate classification method has been detected by our proposed optimized cluster classification method as TPs. Table 4.1 shows the classification performance of various classification architectures at 3 FPs for CRs.

Table 4.1: Classification performance of the CAD systems at 3 FPs for CR using the three classification architectures

<table>
<thead>
<tr>
<th>Type of Classification</th>
<th>Chest Radiographs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Classification</td>
<td>67.5</td>
</tr>
<tr>
<td>Intuitive Cluster Classification</td>
<td>67.7</td>
</tr>
<tr>
<td>Optimized Cluster Classification</td>
<td>75.2</td>
</tr>
</tbody>
</table>

4.4.2. Computed Tomography Results

Potential candidate detector sensitivity of the LIDC dataset used in this research is found to be 87.86%. Among the 280 target nodules, 246 are detected by our candidate detector system. Numerous features for CT scans makes it hard for SFS to choose a good suite of features for holdout classification especially after bifurcation of data. Selection of first feature plays a very vital role in determining good suite of features; a bad choice of first feature might result in poor performance of the CAD system. Hence, to assist SFS and
maintain homogeneity, we seed the first feature for all the SFS processes in CT with *Standard Deviation Separation 3D* feature. This feature has proven to be effective for CAD systems [7]. Now, we repeat the classification processes for CT scans as implemented for CRs and the FROC analysis of 3-fold cross-validation performance on LIDC dataset for all the three classification modes is shown in Figure 4.15. Results clearly suggest that optimized cluster classification outperforms the other methods comprehensively and it follows the trend as observed in CR. A point to note, in [7], final CAD system results are provided after an application of a size filter, whereas all the results are presented here are without the application of this filter. Table 4.2 lists the classification performance of CAD system at 3 FPs for all the classification methods discussed for CT scans.

![FROC comparison of the CAD classification performance for CT scans utilizing different classification architectures](image)

**Figure 4.15:** FROC comparison of the CAD classification performance for CT scans utilizing different classification architectures
Table 4.2: Classification performance of the CAD systems at 3 FPs for CT modality using the three classification architectures

<table>
<thead>
<tr>
<th>Type of Classification</th>
<th>Computed Tomography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Classification</td>
<td>69.9</td>
</tr>
<tr>
<td>Intuitive Cluster Classification</td>
<td>69.9</td>
</tr>
<tr>
<td>Optimized Cluster Classification</td>
<td>74.9</td>
</tr>
</tbody>
</table>
CHAPTER 5

CAD PERFORMANCE AT DIFFERENT THICKNESSES

As mentioned in Chapter 1, classifier performance is compromised when trained on imperfect training data. The imperfections in the training data might be because their statistics may not be representative of the testing data for which the classifier is designed.

CT imagery varies by scanner, slice thickness, slice spacing, reconstruction algorithm, dosage settings and patient. Therefore, imagery from one scanner may not be ideally representative of imagery from another. Some of the feature vectors might be completely different among several databases. A feature that might be very good in one database for distinguishing nodules from non-nodules might not be as effective while testing in another database. This might be because the databases could be statistically different. In addition, a patient’s CT scan might look completely different when scanned using different scanners. These statistical differences can lead to very poor performance of the classifier, which might in turn affect the overall detection of the CAD system. Most CT lung nodule detection classifier systems described in the literature assume the statistics are comparable while training the systems. For the above-mentioned reasons, we focus on CT properties that adversely affect performance of the CAD system. In particular, we focus on slice thickness parameter. Slice thickness is nothing but the thickness of an imaging slice of a
CT scan. As mentioned in Chapter 2, we make use of the diverse LUNA16 dataset for this study.

5.1. Description of the Study

In this chapter, we describe a methodology to study the impact of CAD system performance based on slice thickness of CT scans. We exclusively use 192 CT scans with slice thickness of 1.25mm for this study. CT scans are down-sampled by averaging adjacent slices. We down-sample at different ratios 2, 4 and 8 thereby effectively achieving a simulated thickness of 2.5mm, 5mm and 10mm, respectively. For instance, down-sampling ratio of two is achieved by averaging two consecutive slices in a CT scan and so on. Figure 5.1 illustrates a CT scan from LUNA16 at native thickness of 1.25mm and down-sampled thickness of 2.5mm. Original cue points marked by radiologists at the native slice thickness of 1.25mm are mapped to corresponding equivalent points at different down-sampled thicknesses.

We apply the candidate detector as described in Chapter 3.1 to determine the potential candidates at all thickness stages. We compute a set of 503 features for each candidate. We randomly pick 80 cases with 116 target nodules for testing and rest of the 112 CT scans are utilized for training purposes. We select features solely based on the training dataset using SFS method and classification of the test candidates is performed using a FLD classifier. Note that we utilize the same set of cases for testing and training at all thicknesses. This study at native and various down-sampled stages would help us analyze the CAD performance at different slice thicknesses.
Figure 5.1: Typical CT scan viewed at different slice thicknesses from the LUNA16 dataset at (a) Native: 1.25mm and (b) Down-sampled to 2.5mm
We measure the CAD performance at all thicknesses based on the nodule cue points marked by radiologists at 1.25mm. This would help us compare the performance of CAD system using the same set of nodule cue points at different thicknesses. This approach differs from existing CAD papers where different datasets are utilized at different thicknesses for performance study. Note that we maintain the homogeneity between the train and test cases in terms of slice thickness as emphasis of this experiment is to study the performance of CAD system at different slice thicknesses.

5.2. CAD Performance at Different Slice Thicknesses

In this chapter, we present results for the study presented in Chapter 5.1. Figures 5.2 and 5.3 present images of a small and large nodule at different thicknesses. Both nodule cue points have been transformed to equivalent points at different simulated thicknesses based on radiologists’ marking at 1.25mm slice thickness. Figures 5.2 and 5.3 clearly suggest that nodule tends to lose its shape, size and brightness at higher slice thickness, especially at 10mm. Figures 5.2 and 5.3 also indicate that impact of down-sampling is relatively high for small nodules.

SFS method of feature selection is implemented at the native thickness of 1.25mm and simulated down-sampled stages solely based on their respective training datasets. As mentioned earlier, SFS merit function is measured in terms of AUC from 0-10 FPs. AUC value obtained after selection of each feature is illustrated in Figure 5.4. We choose a point in AUC plot as implemented in [7] to determine the optimal suite of features necessary for
best classification performance. Figure 5.5 presents the FROC comparing the overall CAD performance (including candidate detection and classification) at different thickness levels.
Figure 5.2: Nodule image from the case ‘sub0_p73’ at: (a) native thickness of 1.25mm, (b) simulated down-sampled thickness of 2.5mm, (c) simulated down-sampled thickness of 5mm and (d) simulated down-sampled thickness of 10mm
Figure 5.3: Nodule image from the case ‘sub0_p32’ at: (a) native thickness of 1.25mm, (b) simulated down-sampled thickness of 2.5mm, (c) simulated down-sampled thickness of 5mm and (d) simulated down-sampled thickness of 10mm
Figure 5.4: Selection of classifier features at all thickness stages

Figure 5.5: FROC comparing CAD performance at different thickness levels
Table 5.1 summarizes the overall CAD performance for the testing dataset utilized in this study at different thicknesses with target nodules marked by radiologists at 1.25mm. Note that, every nodule marked by the radiologists at native thickness is transformed into an equivalent cue point at all thickness stages. Table 5.1 indicates that the performance of candidate detector before the application of classifier is consistent at all thickness stages presented in this paper.

**Table 5.1** Overall CAD performance comparison at all thickness levels.

<table>
<thead>
<tr>
<th>Type of Dataset (Based on Thickness)</th>
<th>Candidate detector sensitivity (before classification)</th>
<th>Number of features selected for classification</th>
<th>Overall CAD Performance AUC (0-10 FPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native 1.25mm</td>
<td>91.37</td>
<td>13</td>
<td>7.16</td>
</tr>
<tr>
<td>Simulated 2.5mm</td>
<td>91.37</td>
<td>14</td>
<td>7.29</td>
</tr>
<tr>
<td>Simulated 5mm</td>
<td>92.24</td>
<td>13</td>
<td>6.13</td>
</tr>
<tr>
<td>Simulated 10mm</td>
<td>85.34</td>
<td>14</td>
<td>4.34</td>
</tr>
</tbody>
</table>

Figure 5.6 shows the CAD output from the case ‘sub0_p32’ at all thicknesses. Our CAD system is able detect the nodule at all thicknesses. Figure 5.6 clearly suggests that our CAD system was successfully able to detect a nodule when it is bigger in size and has high contrast. Figure 5.7 shows the histogram distribution of size and contrast of the nodules that were successfully detected at all resolutions respectively.
Figure 5.6: CAD system output for the case ‘sub0_p32’ at different thicknesses (a) Native 1.25mm, (b) Down-sampled from 1.25mm to 2.5mm, (c) Down-sampled from 1.25mm to 5mm and (d) Down-sampled from 1.25mm to 10mm
Figure 5.7: Histogram distribution of (a) size and (b) contrast of the nodules successfully detected by our CAD system at different thicknesses.
In this chapter, we present different training approaches for CAD of lung nodules in CT scans using non-homogenous training data.

6.1. Description of the Methods

As mentioned in Chapter 4, aggregate training method is used in majority of the CAD systems presented in the literature. In this method, we utilize all the training data available at their respective native thickness. In this approach, CT scans are neither re-sampled nor removed thereby using all the available training cases at their respective native thickness.

Homogenous thickness training method utilizes only the cases that match with the thickness and spacing of the testing cases. For instance, if testing is conducted on cases acquired at 1.25mm thickness, then training would be solely based on the data acquired at 1.25mm thereby making it a homogenous thickness classifier.

Finally, we propose a method to maintain the homogeneity between testing and training datasets by re-sampling the entire suite of CT scans to a specific thickness value. This method of classification would help in utilizing all the available training resources and maintaining the homogeneity among the cases (training and testing). We term this approach as common thickness method of classification.
6.2. Experiment Based on 1.25mm Testing Dataset

In this section, we present and compare results for the methodologies proposed in Chapter 6.1 for the testing cases acquired at 1.25mm thickness. We utilize the same set of 80 cases as chosen in Chapter 5 for testing purposes. Rest of the cases available are utilized for training the CAD system. Table 6.1 presents the distribution of the training and testing datasets used for the three different methods of classification.

Table 6.1 Training and testing dataset compositions for different methods of classification – Experiment based on 1.25mm testing dataset.

<table>
<thead>
<tr>
<th>Classification Approach</th>
<th>Training Dataset (Number of cases)</th>
<th>Testing Dataset (Number of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.25mm</td>
<td>2.5mm</td>
</tr>
<tr>
<td>Aggregate</td>
<td>112</td>
<td>283</td>
</tr>
<tr>
<td>Homogenous Thickness</td>
<td>112</td>
<td>0</td>
</tr>
<tr>
<td>Common Thickness</td>
<td>0</td>
<td>283</td>
</tr>
</tbody>
</table>

The candidate detector (before the application of classifier) was successfully able to detect 106 among the 116 target nodules for our testing dataset at both native thickness of 1.25mm and simulated down-sampled thickness of 2.5mm. SFS merit function plot is shown in Figure 6.1. Overall FROC results comparing the various modes of training are
presented in Figure 6.2. Table 6.2 summarizes the overall CAD performance using three different training methods.

Figure 6.1: Selection of classifier features using different training methods for 1.25mm testing dataset
Table 6.2 Overall CAD performance comparison using different training methods for 1.25mm testing dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Candidate detector sensitivity (before classification)</th>
<th>Number of features selected for classification</th>
<th>Overall CAD Performance AUC (0-10 FPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Training</td>
<td>91.37</td>
<td>11</td>
<td>7.36</td>
</tr>
<tr>
<td>Homogenous Thickness Training</td>
<td>91.37</td>
<td>13</td>
<td>7.16</td>
</tr>
<tr>
<td>Common Thickness Training</td>
<td>91.37</td>
<td>14</td>
<td>7.44</td>
</tr>
</tbody>
</table>
Figure 6.2: FROC comparing overall CAD performance using different training methods for 1.25mm testing dataset

6.3. Experiment Based on 2.5mm Testing Dataset

We study and compare the performance of all training methods when testing is conducted on 100 cases acquired at 2.5mm with 114 target nodules. We utilize the rest of the cases available for training purposes. Distribution of training and testing datasets for this experiment is listed in Table 6.3 for the classification methods proposed in Chapter 6.1. Figure 6.3 presents the number of features selected using various classification approaches.
Table 6.3 Training and testing dataset compositions for different methods of classification – Experiment based on 2.5mm testing dataset.

<table>
<thead>
<tr>
<th>Classification Approach</th>
<th>Training Dataset (Number of cases)</th>
<th>Testing Dataset 2.5mm Slice Thickness (Number of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.25mm</td>
<td>2.5mm</td>
</tr>
<tr>
<td>Aggregate</td>
<td>192</td>
<td>183</td>
</tr>
<tr>
<td>Homogenous Thickness</td>
<td>0</td>
<td>183</td>
</tr>
<tr>
<td>Common Thickness</td>
<td>0</td>
<td>183</td>
</tr>
</tbody>
</table>

Figure 6.4 and Table 6.4 present the overall CAD performance for the three different training methods when 100 cases acquired at 2.5mm is utilized for testing. Note that 2.5mm testing cases are not re-sampled for any classification method.
Figure 6.3: Selection of classifier features using different training methods for 2.5mm testing dataset
Table 6.4 Overall CAD performance comparison using different training methods for 2.5mm testing dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Candidate detector sensitivity (before classification)</th>
<th>Number of features selected for classification</th>
<th>Overall CAD Performance AUC (0-10 FPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Training</td>
<td>96.49</td>
<td>13</td>
<td>8.57</td>
</tr>
<tr>
<td>Homogenous Thickness Training</td>
<td>96.49</td>
<td>12</td>
<td>8.68</td>
</tr>
<tr>
<td>Common Thickness Training</td>
<td>96.49</td>
<td>12</td>
<td>8.74</td>
</tr>
</tbody>
</table>
In this section, we present results comparing aggregate and common thickness method of training for a set of 874 cases from the LUNA16 dataset. Figure 6.5 shows the overall FROC using aggregate and common thickness method of training. 14 cases are excluded from the entire LUNA16 dataset (888 cases) due to non-uniform orientation. Overall CAD performance is analyzed based on 10-fold validation. Results clearly indicate that performance is comparable at 2.5mm when compared to the native thickness at a much faster rate.
Figure 6.5: FROC based on 10-fold validation using different training methods for LUNA16 dataset
CHAPTER 7

CONCLUSIONS AND FUTURE WORK

We have presented novel classification processes for improved lung nodule CAD systems in CRs and CT scans. Detailed performance analysis is presented for publicly available datasets thereby setting a benchmark for future research efforts. The optimized cluster classification algorithm performed well on diverse, independent datasets and different modalities thereby demonstrating its versatility. In addition to aggregate classification, we also compare our proposed optimized cluster classification method with intuition-based approach. At 3 FPs, results suggest that performance of aggregate and intuitive cluster classification methods are identical. However, based on overall FROC analysis, intuitive cluster classification method is the least effective. Even though, visually, clustering based on intuitive features appears to be promising, it falters during classification relatively. This might be because the intuition based cluster features are not optimized for classification purposes and there is some uneven distribution of nodules in both the clusters. The aggregate classifier performs reasonably well for both CRs and CT scans as it has been provisioned with all the nodules present in the dataset. Optimized feature selection that are ideal for clustering and classification has never been presented before especially in the field of medical imaging. Applying SFS method for cluster features based on the aggregate classifier features provides a good suite of features that are optimal for
clustering without affecting the classification performance of the CAD system. After clustering based on the optimal suite of cluster features, application of SFS to determine ideal classifier features for each cluster provides the impetus that sets it apart from other classification methods. Overall classification performance of the CAD system using the optimized cluster classification is 75.2% and 74.9% at 3 FPs per case for CR and CT scans thereby outperforming traditional classification methods by about 7.7% and 5.0% respectively. These results are published in the article [46]. Two-cluster division is ideal in this research to get enough distribution of nodules in each cluster. With a larger dataset and more number of diverse nodules, we could cluster the candidates even further.

A valuable area of future work would be to optimize the features for clustering and classification simultaneously. However, this method would be computationally complex, memory demanding and time consuming, especially for CT scans. With the advancement of supercomputers, this could be possible. We could also fuse this optimized method of cluster classification with various other feature selection methods such as genetic algorithm.

Research papers [47-51] provide the initial validation and implementation of deep learning in CAD systems for pulmonary nodule detection and diagnosis. Deep learning approaches for classification of nodules as benign and malignant are provided in [48-51]. However, feature based approaches have outperformed deep learning methods in the literature so far for CAD of lung nodules [47]. With the emergence of deep learning and featureless approaches in various image processing and pattern recognition applications [52,53], applying them effectively for CAD of lung nodules would be a valuable future work.
In addition to feature selection based clustering approach, we have also presented two thickness based studies for CAD of lung nodules in CT scans. First study presented the performance of CAD system at various thickness levels. FROC results presented in Figure 5.5 and Table 5.1 indicate that the CAD system provides comparable performance at native thickness and simulated down-sampled thickness of 2.5mm. In fact, CAD system achieves good performance at a much faster rate (2x) with reduced memory consumption when down-sampled to a simulated thickness of 2.5mm. However, classification performance deteriorates considerably when down-sampled further than 2.5mm. Our experimental results suggest that with same amount of data across various thickness values (1.25mm, 2.5mm, 5mm and 10mm), 2.5mm is the most effective in terms of accuracy, dosage level, computation and memory consumption.

Later, we presented results comparing CAD performance using three training methods for non-homogenous data. Figures 6.2, 6.4 and 6.5 indicate that common thickness method of training (at 2.5mm) provides the best results for testing data acquired at both 1.25mm and 2.5mm thickness. Common thickness method helps in maintaining the homogeneity among the cases (training and testing) and in utilizing all the cases available for training. This performance is closely followed by aggregate method of training albeit using increased memory and more computation time. Homogenous thickness method of training could be utilized when there are sufficient training cases that match with the thickness of the testing cases.

A valuable area of future work would be to study the performance of CAD systems at various dosage levels. A study that aids in determining optimal dosage level for best CAD performance would be of great benefit for lung cancer screening.


