University of Central Florida
STARS

Electronic Theses and Dissertations | Doctoral Dissertation (Open Access)

Learning Algorithms for Fat Quantification and Tumor Characterization

2018

Sarfaraz Hussein
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Computer Sciences Commons

STARS Citation

Hussein, Sarfaraz, "Learning Algorithms for Fat Quantification and Tumor Characterization" (2018). *Electronic Theses and Dissertations*. 5962.
https://stars.library.ucf.edu/etd/5962

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.

University of Central Florida
STARS
Showcase of Text, Archives, Research & Scholarship

LEARNING ALGORITHMS FOR FAT QUANTIFICATION AND TUMOR CHARACTERIZATION

by

SARFARAZ HUSSEIN
B.E. NED University of Engineering and Technology, Karachi, 2011

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2018

Major Professor: Ulas Bagci

© 2018 Sarfaraz Hussein

ABSTRACT

Obesity is one of the most prevalent health conditions. About 30% of the world's and over 70% of the United States' adult populations are either overweight or obese, causing an increased risk for cardiovascular diseases, diabetes, and certain types of cancer. Among all cancers, lung cancer is the leading cause of death, whereas pancreatic cancer has the poorest prognosis among all major cancers. Early diagnosis of these cancers can save lives. This dissertation contributes towards the development of computer-aided diagnosis tools in order to aid clinicians in establishing the quantitative relationship between obesity and cancers. With respect to obesity and metabolism, in the first part of the dissertation, we specifically focus on the segmentation and quantification of white and brown adipose tissue. For cancer diagnosis, we perform analysis on two important cases: lung cancer and Intraductal Papillary Mucinous Neoplasm (IPMN), a precursor to pancreatic cancer. This dissertation proposes an automatic body region detection method trained with only a single example. Then a new fat quantification approach is proposed which is based on geometric and appearance characteristics. For the segmentation of brown fat, a PET-guided CT co-segmentation method is presented. With different variants of Convolutional Neural Networks (CNN), supervised learning strategies are proposed for the automatic diagnosis of lung nodules and IPMN. In order to address the unavailability of a large number of labeled examples required for training, unsupervised learning approaches for cancer diagnosis without explicit labeling are proposed. We evaluate our proposed approaches (both supervised and unsupervised) on two different tumor diagnosis challenges: lung and pancreas with 1018 CT and 171 MRI scans respectively. The proposed segmentation, quantification and diagnosis approaches explore the important adiposity-cancer association and help pave the way towards improved diagnostic decision making in routine clinical practice.

EXTENDED ABSTRACT

This dissertation explores the learning algorithms that can be used to extract imaging markers reliably so that the obesity-cancer relationship can be studied. The novel learning algorithms developed in this dissertation are general and can be modified for other applications similar to those examined here. More specifically, we first investigate the automatic detection of white and brown adipose tissues using Positron Emission Tomography/Computed Tomography (PET/CT) scans and develop methods for the quantification of these tissues at the whole-body and body-region levels. We propose a patient-specific automatic adiposity analysis system with two modules i.e. detection, segmentation and quantification of white and brown adipose tissue. In the first part, we detect white adipose tissue (WAT) and its two sub-types from CT scans, Visceral Adipose Tissue (VAT) and Subcutaneous Adipose Tissue (SAT). This process relies conventionally on manual or semi-automated segmentation, leading to inefficient solutions. Our novel framework addresses this challenge by proposing an unsupervised learning method to separate VAT from SAT in the abdominal region for the clinical quantification of central obesity. This step is followed by a context-driven label fusion algorithm through sparse 3D Conditional Random Fields (CRF) for volumetric adiposity analysis. Evaluations were performed on 151 CT scans and our proposed approach outperformed other methods by a significant margin.

In the second part of the dissertation, we automatically detect, segment, and quantify brown adipose tissue (BAT) using PET scans because unlike WAT, BAT is metabolically active. After identifying BAT regions using PET, we perform a co-segmentation procedure utilizing asymmetric complementary information from PET and CT. Finally, we present a new probabilistic distance metric for differentiating BAT from non-BAT regions. In both parts, we perform automatic body-region detection using one-shot learning. Experimental evaluations conducted on 111 PET/CT scans achieve state-of-the-art performances in brown adiposity quantification.

In the third part, we propose different supervised strategies for the characterization of lung nodules and Intraductal Papillary Mucinous Neoplasms (IPMN). Risk stratification of lung nodules as benign or malignant is one of the most important tasks in lung cancer diagnosis, staging and treatment planning. While the variation in the appearance of the nodules remains large, there is a need for a fast and robust computer-aided system. We propose a new end-to-end trainable multi-view deep Convolutional Neural Network (CNN) for nodule characterization. The trained network is used to extract features from the input image followed by a Gaussian Process (GP) regression to obtain the malignancy score. We also empirically establish the significance of different high-level nodule attributes such as calcification, sphericity and others for malignancy determination. These attributes are found to be complementary to the deep multi-view CNN features and a significant improvement over other methods is obtained.

We next propose a 3D Convolutional Neural Network (CNN) based nodule characterization strategy. With a completely 3D approach, we utilize the volumetric information from a CT scan which would be otherwise lost in the conventional 2D CNN based approaches. In order to mitigate the need for a large amount of training data for CNN, we resort to transfer learning to obtain highly discriminative features. Moreover, we also acquire a task-dependent feature representation for six high-level nodule attributes and fuse this complementary information via a Multi-task learning (MTL) framework.

For the diagnosis of IPMN, we propose a Convolutional Neural Network (CNN) based computer-aided diagnosis (CAD) system by utilizing multi-modal Magnetic Resonance Imaging (MRI). In our proposed approach, we use minimum and maximum intensity projections to ease the annotation variations among different slices and type of MRIs. Then, we present a CNN to obtain deep feature representation corresponding to each MRI modality (T1-weighted and T2-weighted). At the final step, we employ canonical correlation analysis (CCA) to perform a fusion operation at the feature level, leading to discriminative canonical correlation features. Extracted features are used for

classification. Our results indicate significant improvements over other potential approaches to solve this important problem. The proposed approach doesn't require explicit sample balancing in cases of imbalance between positive and negative examples. To the best of our knowledge, our study is the first to automatically diagnose IPMN using multi-modal MRI.

In medical imaging, the availability of labeled data warrants extensive commitments of time and efforts from radiologists. Hence, developing automated diagnostic systems with a minimum amount of training data is important. In the final part of this dissertation, we propose an unsupervised machine learning approach to perform tumor classification. First, we cluster the data into groups and then improve the tumor classification using $\propto$SVM which is learned on the label proportions. We also explore the contribution of features extracted from different deep learning architectures in improving the classification. Furthermore, we study the effect of supervised learning approaches on the classification performance. Finally, we analyze unsupervised representation learning with Generative Adversarial Networks (GAN) for tumor classification. We evaluate our proposed approaches for the diagnosis of lung nodules and IPMN, and obtain improved classification performance than other methods.

To my mother

~

To my father

and all family

for their unconditional love,

efforts and sacrifices

~

To every investor of time,

effort and resources

in my learning

~

To all pursuers of

knowledge and wisdom

ACKNOWLEDGMENTS

All praises are due to Allah, for countless gifts and blessings, who is the most gracious and the most merciful. I would like to thank my advisor Dr. Ulas Bagci for his efforts, guidance, and support throughout my graduate studies. I have been fortunate enough to learn not only about research but about different aspects of life which include but are not limited to commitment, discipline, and hard work. He kept me motivated during difficult phases of this journey. I'm also especially grateful to Dr. Mubarak Shah for the opportunity of working with him during the first few years of my graduate studies and for important lessons and guidelines about research. I'm very grateful to Dr. Mark Heinrich and Dr. Marianna Pensky for serving on my committee.

I would also like to thank our collaborators at the Mayo Clinic Jacksonville, namely, Dr. Michael Wallace, Dr. Juan Corrall, Dr. Pujan Kendal and Dr. Candice Bolan, as well as our collaborators at the National Institute of Health (NIH).

I would like to acknowledge the financial support for this work from Mayo-UCF Seed Funding as well as Zika Research Grant from the Florida Department of Health.

It is said that it takes a village to raise a child. This is true for a Ph.D. as well. I would like to thank all my teachers that kindled the love of knowledge in me. I would like to acknowledge Professor Shahab Tehzeeb for motivating me and guiding me towards graduate studies. I would like to thank all past and present members of CRCV including Dr. Imran Saleemi, Dr. Waqas Sultani, Dr. Khurram Soomro, Dr. Dong Zhang, Dr. Yang Yang, Mahdi Kalayeh, Harish RaviPrakash, Shervin Ardeshir, Muhammad Abdullah Jamal, Amir Mazaheri, Aliasghar Mortazi, Cherry Place, Brittany Kaval and Tonya LaPrarie for their support and good memories. I would like to acknowledge the support of all my friends, including Kumail Abbas, Shabeer Kirmani, Saad Khan, Fahd Khan, Asad Waraich, Mehboob-ur-Rehman and Haider Ali.

I would also like to thank Vivek Singh and Terrence Chen from Siemens Healthineers as well as Kunlin Cao and Qi Song from CuraCloud Corporation for their support during my internships.

Last but most important, I would like to acknowledge the unconditional love and efforts of my parents, appa, brothers, sisters in law, nieces and nephews.

TABLE OF CONTENTS

LIST OF FIGURES xv

LIST OF TABLES xxii

CHAPTER 1: INTRODUCTION 1

1.1 Segmentation and Quantification of White Adipose Tissue 2

1.2 Segmentation and Detection of Brown Adipose Tissue 6

1.3 Lung Nodule Diagnosis with Convolutional Neural Networks 6

1.4 Diagnosis of IPMN with Multi-Modal Imaging 8

1.5 Unsupervised Learning Approaches for Tumor Characterization 9

1.6 Dissertation Organization 10

CHAPTER 2: LITERATURE REVIEW 11

2.1 Fat Segmentation and Quantification 11

2.2 Malignancy Characterization of Lung Nodules 12

2.3 Diagnosis of IPMN 14

2.4 Unsupervised Learning Approaches 15

CHAPTER 3: SEGMENTATION AND QUANTIFICATION OF WHITE ADIPOSE TISSUE 17

3.1 Region Detection in whole-body CT volumes 17

3.2 SAT-VAT Separation and Quantification 19

3.3 Results 25

3.3.1 Data for Central Obesity Quantification: 25

3.3.2 Parameters and evaluations metrics: 26

3.3.3 Comparisons: 27

3.3.4 Computation Time: 28

3.4 Summary 29

CHAPTER 4: BROWN FAT DETECTION AND SEGMENTATION 30

4.1 Method for BAT Detection and Segmentation 31

4.2 Results 36

4.2.1 Data for Quantification of BAT: 36

4.2.2 Evaluation of Head-neck and Thorax Region Detection: 38

4.2.3 Evaluation of BAT Delineation: 39

4.2.4 Comparisons with Other Approaches: 39

4.2.5 Evaluation of BAT Region Proposals: 40

4.3 Summary 40

CHAPTER 5: SUPERVISED LEARNING FOR TUMOR CHARACTERIZATION 44

5.1 TumorNET: Lung Nodule Characterization Using Multi-View 2D Convolutional Neural Network with Gaussian Process 45

5.1.1 Contributions: 45

5.1.2 Materials 46

5.1.3 Method 47

5.1.3.1 Multiview Deep Convolutional Neural Network: 47

5.1.3.2 Gaussian Process (GP) Regression: 48

5.1.4 Experiments and Results 49

5.1.5 Summary 51

5.2 Malignancy characterization of Lung Nodules using 3D CNN Based Multi-Task Learning: 53

5.2.1 Method 55

5.2.1.1 Problem Formulation 55

5.2.1.2 Network Architecture and Transfer Learning 56

5.2.1.3 Multi-task learning 57

5.2.1.4 Optimization 59

5.2.2 Experiments 60

5.2.2.1 Data 60

5.2.2.2 Results 60

5.2.3 Summary 63

5.3 Deep Multi-modal Diagnosis of IPMN 65

5.3.1 CNN for Multi-modal Feature Representation: 65

5.3.2 Feature Fusion with Canonical Correlation Analysis: 67

5.3.3 Materials 69

5.3.4 Experiments and Results 70

5.3.5 Summary 73

CHAPTER 6: UNSUPERVISED TUMOR CHARACTERIZATION 75

6.1 Initial Label Estimation 76

6.2 Learning with the Estimated Labels 77

6.3 Calculating Label Proportions 78

6.4 Experiments and Results: 79

6.4.1 Are Deep Features good for Unsupervised Classification? 81

6.4.2 Classification using Supervised Learning 83

6.4.3 Generative Adversarial Networks (GAN) 84

6.5 Summary 86

CHAPTER 7: CONCLUSION AND FUTURE DIRECTIONS 88

7.1 Obesity 88

7.2 Tumor Characterization 89

7.3 Future Directions 90

LIST OF REFERENCES 93

LIST OF FIGURES

1.1 The organization of the dissertation is shown. The clinical motivation is to help clinicians in establishing the quantitative relationship between obesity and different cancers. Image segmentation, object detection and quantification of white and brown adipose tissue drive the obesity and metabolism part of the study, whereas lung and pancreatic cancers are our focus for cancer diagnostic system. We focus on non-invasive quantification methods through radiology images. 3

1.2 An illustration of different types of adipose tissues in Positron Emission Tomography (PET) and Computed Tomography (CT) scans. (A) signifies the difference at the cellular level between Brown Adipose Tissue (BAT) and White Adipose Tissue (WAT). In contrast to WAT, BAT is metabolically active and consumes energy. (B) shows Subcutaneous Adipose Tissue (SAT) and Visceral Adipose Tissue (VAT) in a coronal view of CT. The red boundary illustrates the thin muscular wall separating these two sub-types. The wall remains mostly discontinuous, making SAT-VAT separation significantly challenging. (C) depicts metabolically active BAT in PET (left/middle) and PET/CT fusion (right). 4

3.1 An overview of the proposed SAT-VAT separation method. Once the abdominal region is detected, Total Adipose Tissue (TAT) is segmented using the CT intensity interval known for fat tissue. Initial Subcutaneous-Visceral adipose tissue boundary is estimated by evaluating multiple hypothesis points. Geometric Median Absolute Deviation (MAD) and appearance based Local Outlier Scores (LoOS) are then combined within the 3D Conditional Random Field (CRF) based label fusion. 19

3.2 An illustration of skin boundary and hypothesis points along the radii connecting S with its centroid C. For each point (yellow) in S, a set of hypotheses (blue) is generated which is along the line connecting the skin boundary point with the centroid C. 21

3.3 t-SNE visualizations using (a) Euclidean and (b) Normalized Correlation distances. Better separation of classes (red and blue) can be clearly seen in (b). 23

3.4 Visceral Adipose Tissue (red) and Subcutaneous Adipose Tissue (green) segmentations are illustrated for two subjects (one with BMI< 25, another with BMI> 30) at the chosen abdominal slice level along with their volume renderings. Several abdominal slices are also shown for central adiposity accumulation. 27

4.1 An overview of the proposed Brown Adipose Tissue (BAT) detection and segmentation system. Given the head-neck and thorax regions, adipose tissue is identified using CT thresholding intervals (Step 1). Using the corresponding PET scans, segmentation seeds are sampled in accordance with high uptake regions (Step 2). PET-CT co-segmentation is performed using Random Walk (Step 3) followed by false positive removal (Step 4) using Total Variation and Cramér-von Mises distances. 31

4.2 For three different anatomical levels (columns), row (A) shows reference standards (white); row (B) demonstrates the results from CT thresholding where pink (inner) and blue (outer) contours show brown fat delineation (blue contour shows fat region near skin boundary which leaks into the body cavity and also overlaps with pink contour as in the first column); row (C) comprises the results from ROI (Region of Interest) based CT thresholding, where orange boxes show user drawn ROIs and blue contours are the brown fat delineation results; row (D) shows the results from conventional PET thresholding, where green contours show output BAT delineations; row (E) depicts the ROI based PET thresholding; and row (F) demonstrates the proposed algorithm's delineation results using PET and CT jointly. (G) Dice Similarity Coefficients (DSC) of the proposed method in comparison with ROI based PET thresholding, PET thresholding, ROI based CT thresholding, and CT thresholding methods are shown. 38

5.1 An overview of the proposed method. First, the median intensity projection is performed across each axis, followed by their concatenation as three channels of an image. Data augmentation is performed using scaling, rotation, adding Gaussian, Poisson, Salt and Pepper (S&P) and Speckle Noise. A CNN with 5 convolution and 3 fully connected layers is trained from scratch. For testing, the 3 channel image is passed through the trained network to get a 4096-dimensional feature from the first fully connected layer. Finally, the malignancy score is obtained using the Gaussian Process regression. 46

5.2 Qualitative results showing median intensity projected images for correctly (green) and incorrectly (red) scored lung nodules. (a) and (b) show malignant and benign nodules respectively where each row shows different cases and column represents different views (axial, sagittal, coronal). 52

5.3 Lung nodule attributes with different scores. As we move from the top (attribute missing) to the bottom (attribute with the highest prominence), the prominence of the attributes increases. Column (a) and (b) show calcified and spherical nodules; (c) represents margin where the top row is for poorly defined nodules and the bottom row shows well-defined nodules. Column (d) and (e) show lobulated and spiculated nodules whereas (f) represents nodules with different textures. The top row in (f) represents a non-solid nodule and the bottom row shows a solid nodule. The graph in (g) shows the number of nodules with different malignancy scores 54

5.4 An overview of the proposed approach. First, we fine-tune 3D CNNs using labels for malignancy and six attributes. Given the input volume, we pass it through different 3D CNNs each corresponding to an attribute (task). The network comprises 5 convolution, 5 max pooling, and 2 fully connected layers. We use the output from the first fully connected layer as the feature representation. The features from different CNNs are fused together using graph regularized sparse least square optimization function to obtain coefficient vectors corresponding to each task. During the testing phase, we multiply the feature representation of the testing image with the coefficient vector to obtain the malignancy score. 55

5.5 Plots to show classification accuracy against various threshold values for the average absolute score difference. The graphs in (a) and (b) represent results from 10 different cross-validation (CV) sets. It can be seen that the classification accuracy increases, as we increase the threshold value for the absolute score difference. The graph (c) shows the improved performance of the proposed method in comparison with GIST+LASSO and 3DCNN with Trace Norm. 62

5.6 Qualitative results using our proposed approach. (a) and (b) show axial views of benign and malignant nodules respectively, where first four columns consist of successful cases (where prediction was within ± 1 of the expert score) and the last column (after dotted line) shows failure cases. 63

5.7 Effect of fine-tuning on 3D CNN features. *t*-SNE visualization for features obtained from (a) pre-trained network and (b) network after fine-tuning. The separation between features belonging to two classes, i.e. benign nodules (represented in blue) and malignant nodules (shown in red) can be readily observed in (b). 64

5.8 An overview of the proposed method. First, the minimum and maximum intensity projections are computed corresponding to T1 and T2 scans respectively. The intensity projections are then fed into a pre-trained Convolutional Neural Network (CNN) to obtain feature representation. Canonical Correlation Analysis (CCA) based feature fusion is performed in order to obtain discriminative and transformed feature representation. Finally, an SVM based classifier is employed to obtain the final label (normal or IPMN). 67

5.9 An illustration to show CCA computed between features from two different imaging modalities i.e. T1 and T2 MRI scans 68

5.10 Qualitative results of our proposed approach, showing minimum and maximum intensity projected images for T1 and T2 scans on left and right respectively. Each row represents a different case where the images correctly classified as IPMN are shown in green, whereas the misclassification of IPMN as normal are shown in red. 72

6.1 An outline of the proposed unsupervised approach. Given the input images, we compute GIST features and perform k-means clustering to get the initial set of labels which can be noisy. Using the set of labels, we compute label proportions corresponding to each cluster/group (Eq. 6.4). We finally employ $\propto$SVM to learn a discriminative model using the features and label proportions. 76

6.2 Representation of lung nodules in (a) axial CT slices and in (b) volume renderings. The top row shows malignant nodule where the second row depicts an instance of benign nodule. The nodules are highlighted in red in (a) and are pointed by arrows in (b). 79

6.3 Axial T2 MRI scans illustrating pancreas. The top row shows different ROIs of pancreas, along with a magnified view of a normal pancreas (outlined in blue). The bottom row shows ROIs from subjects with IPMN in the pancreas, which is outlined in red. 80

6.4 Influence of deep learning features obtained from different layers of a VGG network with and without ReLU non-linearities. The graph on the left shows accuracy, sensitivity and specificity for unsupervised lung nodule classification (clustering), whereas the right one shows the corresponding results for IPMN. 82

6.5 The generator architecture of DCGAN used to perform unsupervised representation learning. 85

LIST OF TABLES

3.1 Abdominal region detection results measured by Intersection over Union (higher the better) and average absolute slice difference (lower the better) along with standard error of the mean (SEM). The IoUs are in 0-1 interval. . . 26

3.2 Segmentation and quantification results for Subcutaneous Adipose Tissue (SAT) and Visceral Adipose Tissue (VAT) evaluated by Dice Similarity Coefficient (higher the better) and Mean Absolute Error (lower the better) along with standard error of the mean (SEM). 28

4.1 Head-neck and Thorax Region detection results measured by Intersection over Union (IoU) and average absolute slice difference along with standard error of the mean (SEM). IoUs are in 0-1 interval. 40

5.1 Comparison of the proposed approach with Support Vector Regression, Elastic Net and LASSO using accuracy measure and standard error of the mean (SEM). 50

5.2 Regression accuracy and standard error (SEM) using the combination of high-level attributes and CNN features. 51

5.3 Classification accuracy and mean absolute score difference (lower the better) of the proposed multi-task learning method in comparison with the other methods. 61

5.4 Results for accuracy, sensitivity and specificity of the proposed multi-modal fusion approach along with standard error of the mean (SEM) in comparison with single modality and feature concatenation based approaches. 70

5.5 Classification accuracy along with standard error of the mean (SEM) for three class classification (normal, low grade IPMN and high grade IPMN) of proposed approach in comparison with other approaches. 71

6.1 Average classification accuracy, sensitivity and specificity of the proposed *unsupervised* approach for IPMN and lung nodule classification with other methods 81

6.2 Classification of IPMN and Lung Nodules using different features and supervised learning classifiers. 83

6.3 Supervised classification for IPMN and Lung Nodules using unsupervised feature learning via Deep Convolutional Generative Adversarial Networks (DC-GAN) 86

CHAPTER 1: INTRODUCTION

The idea of automatically detecting and classifying objects with high accuracy, robustness and efficiency has long enticed researchers. Several attempts have been made to model human vision capabilities into machines so as to serve this purpose. There has been an increased interest in performing automatic and computer-aided detection and diagnosis (CAD). Conventionally, radiologists manually analyze images to detect and diagnose clinical anomalies which can then be referred for biopsies. As medical imaging devices are becoming more and more accessible to the general population, the number of images to analyze is drastically increasing. This calls for an increased manpower (radiologists) available to analyze them. However, there is one radiologist available for 10,000 people in the United States [1]. This number soars to 100,000 for India [2] and more than 200,000 for Kenya [3]. This increases the workload of radiologists where an average radiologist analyzing MRI or CT should study one image every 3-4 seconds in an 8-hour workday to satisfy workload requirements [4]. The heavy workload leads to an increase in the number of diagnostic errors where the retrospective error rate among radiologic examinations is around 30% with real-time errors averaging around 3-5% [5]. Annually misdiagnosis results in an estimated number of 40,000 to 80,000 US hospital deaths [6].

The CAD tools can assist radiologists by providing them a second opinion for better diagnosis. Moreover, these automated tools can efficiently sift through images to mark cases which require immediate attention. The algorithms perform detection and diagnosis much faster than a human and are especially helpful where significant exertion is required such as segmentation tasks. Moreover, these approaches are less prone to biases which are quite common among radiologists. The common biases in medical reasoning are analyzed in [7]. Although there are CAD tools developed to detect and diagnose different pathologies, they are not accurate and robust enough to be employed in clinics. These CAD systems give a higher number of false positives thus making

them inefficient for clinical applications. Moreover, there is no CAD system to help analyze the obesity-cancer association. This dissertation is an effort to bridge this gap.

The overarching goal of this dissertation is to facilitate clinicians in establishing the quantitative relationship between obesity/metabolism and cancers. In the clinical literature, the relationship between obesity and different types of cancers is thoroughly discussed, however, the use of CAD systems can help automate the process and make the study scalable to larger cohorts. Specifically, we focus on two cancer types, lung and pancreatic cancers, that are not investigated in the literature extensively in terms of metabolic health.

In order to represent metabolic information, one needs to extract fat volumetry and depot (location) in the body. In this dissertation, CAD approaches for the segmentation, detection, and quantification of white and brown fat are proposed. Moreover, methods for the diagnosis of lung cancer and IPMN which is the precursor to the pancreatic cancer are presented. The overall organization of the dissertation along with the clinical motivation, imaging modalities, and technical objectives are shown in Figure 1.1. In terms of technical contributions, we develop algorithms to extract cancer and metabolism related markers and perform their quantification and automatic characterization.

1.1 Segmentation and Quantification of White Adipose Tissue

Brown adipose tissue (BAT), *brown fat*, and white adipose tissue (WAT) are the two types of adipose tissues found in mammals (Figure 1.2-A). Quantification of WAT and its subtypes is an important task in the clinical evaluation of obesity, cardiac diseases, diabetes, and other metabolic syndromes [8–10]. Among them, obesity is one of the most prevalent health conditions.

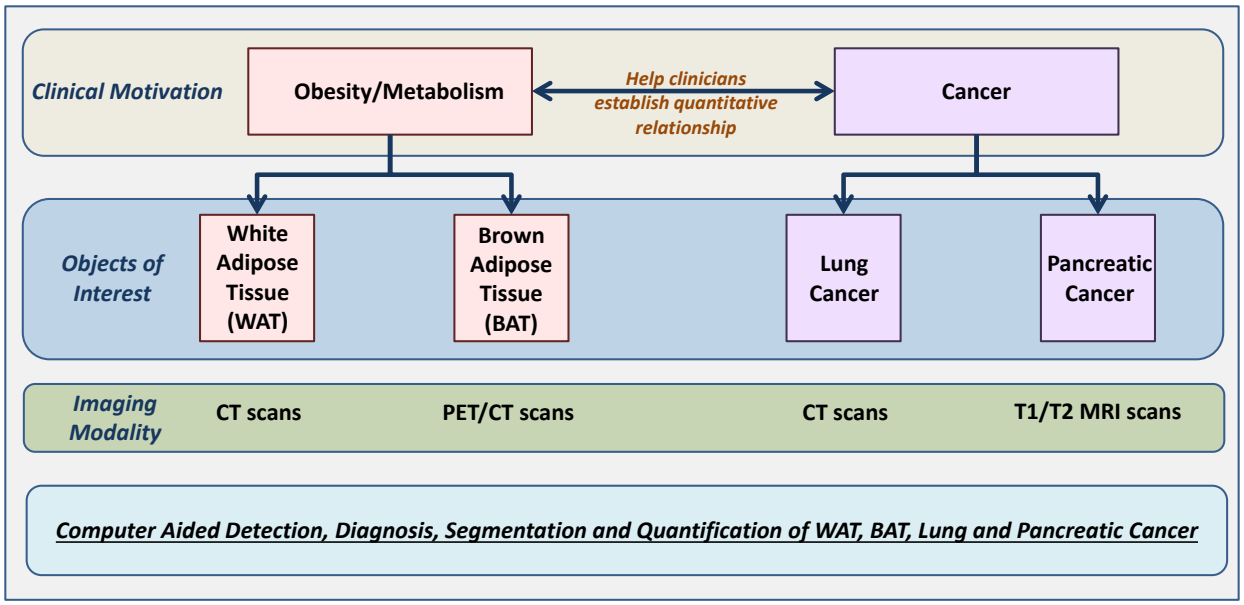


Figure 1.1: The organization of the dissertation is shown. The clinical motivation is to help clinicians in establishing the quantitative relationship between obesity and different cancers. Image segmentation, object detection and quantification of white and brown adipose tissue drive the obesity and metabolism part of the study, whereas lung and pancreatic cancers are our focus for cancer diagnostic system. We focus on non-invasive quantification methods through radiology images.

About 30% of the world's and over 70% of the United States' adult populations are either overweight or obese [11,12], causing an increased risk for cardiovascular diseases, diabetes, and certain types of cancer. Central obesity (also known as abdominal obesity) is the excessive buildup of fat in the abdominal region. Traditionally, Body Mass Index (BMI) has been used as a measure of obesity and metabolic health; however, BMI remains inconsistent across subjects, especially for underweight and obese individuals. Instead, volumetry of the abdominal fat, i.e., Visceral Adipose Tissue (VAT), is considered as a reliable, accurate, and consistent measure of *body fat distribution*. As VAT manifests itself mainly in the abdominal region, it is regarded as an important marker for evaluating central obesity. In clinical literature, the association between VAT and different diseases has been studied extensively.

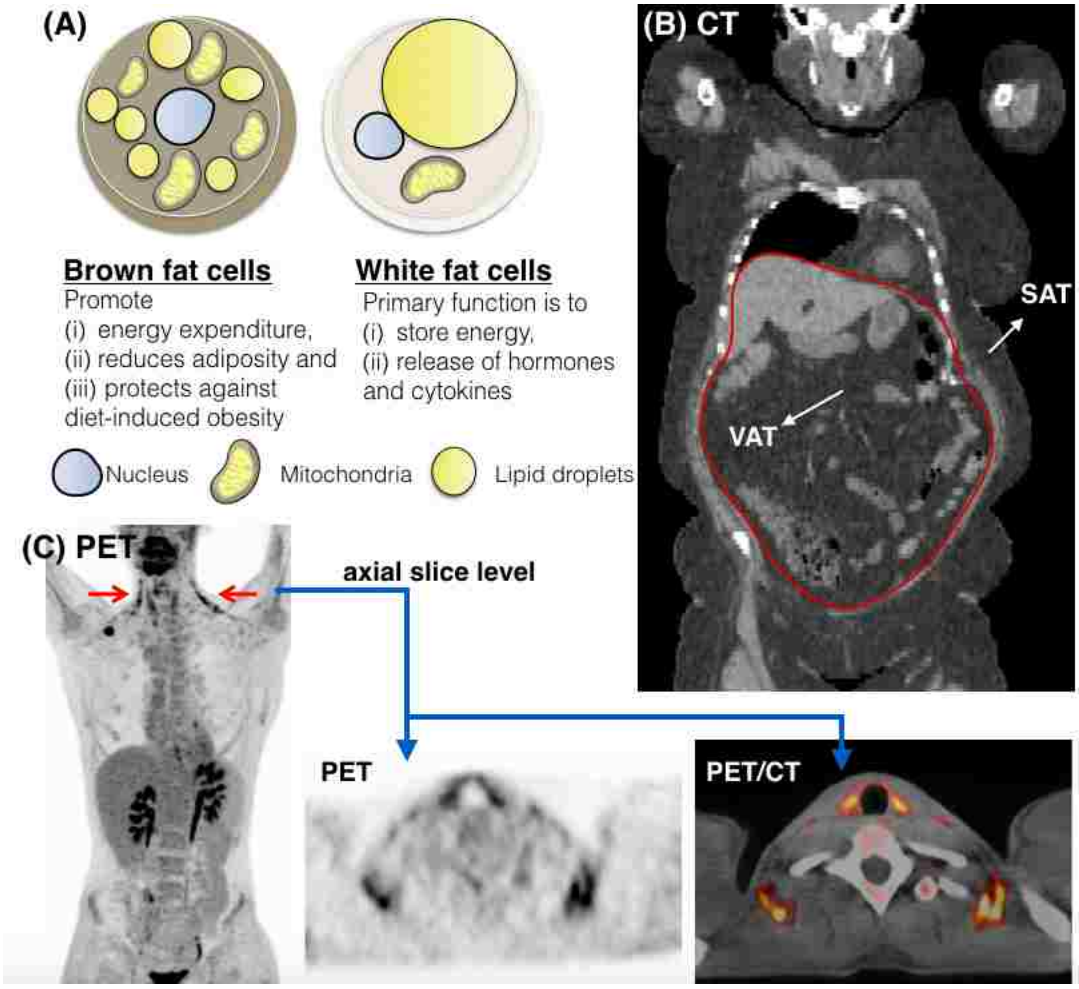


Figure 1.2: An illustration of different types of adipose tissues in Positron Emission Tomography (PET) and Computed Tomography (CT) scans. (A) signifies the difference at the cellular level between Brown Adipose Tissue (BAT) and White Adipose Tissue (WAT). In contrast to WAT, BAT is metabolically active and consumes energy. (B) shows Subcutaneous Adipose Tissue (SAT) and Visceral Adipose Tissue (VAT) in a coronal view of CT. The red boundary illustrates the thin muscular wall separating these two sub-types. The wall remains mostly discontinuous, making SAT-VAT separation significantly challenging. (C) depicts metabolically active BAT in PET (left/middle) and PET/CT fusion (right).

For instance, visceral obesity quantified through Computed Tomography (CT) was found to be a significant risk factor for prostate cancer [13]. In [14], visceral adiposity was found to be a significant predictor of disease-free survival rate in resectable colorectal cancer patients. In contrast

to Subcutaneous Adipose Tissue (SAT), VAT was concluded to have an association with incident cardiovascular disease and cancer after adjustment for clinical risk factors and general obesity [15]. Speliotes et al. [16] found VAT as the strongest correlate of fatty liver among all the other factors used in their study. In [17], VAT was found to be an independent predictor of all-cause mortality in men after adjustment for abdominal subcutaneous and liver fat. This clinical evidence shows that the robust and accurate quantification of VAT can help improve the identification of risk factors, prognosis, and long-term health outcomes.

However, automatic separation of VAT from SAT in CT images is not trivial because both VAT and SAT regions share similar intensity characteristics (Hounsfield unit (HU)), and are vastly connected (Figure 1.2-B). To segregate these two tissues, radiologists usually use various morphological operations along with manual interactions, but this process is subjective and unattractive for routine clinical evaluations. A set of representative slices at or near the umbilical level is often used for quantifying central obesity [18]. Still, these selections do not accurately infer volumetric quantification. Thus, inefficient and inaccurate quantification remains a major obstacle in the clinical evaluation of body fat distribution.

Our contributions: To the best of our knowledge, the proposed system is the first fully automated method for detecting, segmenting, and quantifying SAT and VAT regions from radiology scans (CT images). First, we propose an automated abdominal and thorax region detection algorithm, based on deep learning features. Second, we develop an *unsupervised* learning method for separating VAT from SAT using appearance (via Local Outlier Scores) and geometric (via Median Absolute Deviation) cues. For volumetric quantification, we integrate contextual information via a sparse 3D Conditional Random Fields (CRF) based label fusion algorithm. Our work can be considered as the largest central obesity quantification study (151 CT scans) to date, validating accurate region and abdominal fat detection algorithms.

1.2 Segmentation and Detection of Brown Adipose Tissue

BAT is important for thermogenesis and is considered as a natural defense against hypothermia and obesity [19]. Since BAT is metabolically active, the sensitivity of Positron Emission Tomography (PET) imaging to detect BAT regions is much higher than that of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) (Figure 1.2-C). However, PET lacks specificity due to limited structural information. When combined with CT and/or MRI, both specificity and sensitivity are increased due to the inclusion of anatomical sites into the evaluation framework. Despite rapid improvements in the imaging facets of BAT detection, the available methods are limited to manual and semi-automated strategies; hence, they are time-consuming and non-reproducible.

Our contributions: For our contributions in BAT detection and segmentation, we first use a fixed HU interval to identify total adipose tissue (TAT) from CT images. Next, we devise a seed sampling scheme for extracting foreground and background cues from high uptake regions of PET images in head-neck and thorax regions only. The identified seeds are propagated into the corresponding CT scans as well. This is followed by a PET-guided image co-segmentation on the hypergraph (PET/CT) to delineate potential BAT regions. Lastly, for false positive (FP) rejection, we propose a novel probabilistic metric that combines *Total Variation* and *Cramér-von Mises* distances to differentiate BAT regions from non-BAT regions.

1.3 Lung Nodule Diagnosis with Convolutional Neural Networks

Cancer is among the leading causes of death worldwide. Out of 8.2 million deaths due to cancer worldwide, lung cancer accounts for the highest number of mortalities i.e. 1.59 million [20]. Moreover, lung cancer still has a 5-year survival rate of 17.8% where only 15% of all cases are diagnosed at an early stage [21]. Risk stratification of lung nodules can aid in identifying cancer

stage leading to improved treatment and higher chances of survival. In addition, any significant development to accurately and automatically characterize lung nodules can save significant manual exertion as well as valuable time.

Early diagnosis is one of the ways to reduce deaths related to lung cancer [22]. In this regard, lung screening programs are especially beneficial. Low Dose CT scans are usually used to perform lung nodule diagnosis, including both detection and risk stratification. Although CT imaging remains the gold standard for lung cancer detection and diagnosis, Computer-Aided Diagnosis (CAD) and quantification tools are often necessary. Moreover, research in developing CAD algorithms can help explore the domain of imaging features and biomarkers which can be then studied by radiologists to further improve clinical decision making.

The development of a fast, robust and accurate system to perform risk stratification of lung nodules is therefore of significant importance. Specifically, the availability of large publicly available datasets such as LIDC-IDRI from Lung Image Database Consortium [23] has helped accelerate the research in this regard. However, the variability in nodule characteristics, including shape, size, intensity, location, as well as uncertainty among radiologists' interpretations have made this problem particularly challenging. The advancement in machine learning methods, including the development of novel classification and feature learning techniques, has increased the efficacy of this task. However, there remains a substantial progress to be done in order to develop a CAD system attractive enough to be used in routine clinical evaluations of lung nodules.

Our Contributions:

- We utilize the full potential of a deep neural network by training a 2D network from scratch on lung nodule dataset. In order to generate enough examples for training, we perform data augmentation. Finally the malignancy score is predicted with Gaussian Process Regression.

- We analyze the influence of high level nodule attributes such as calcification, lobulation, spiculation and others for malignancy prediction. Our proposed approach is based on 3D CNN network with multi-task learning to jointly learn the relationship between attributes and malignancy.
- We evaluate our proposed approaches on one of the largest publicly available datasets consisting of more than 1000 low-dose CT scans.

1.4 Diagnosis of IPMN with Multi-Modal Imaging

Among all cancers, pancreatic cancer has the poorest prognosis with a 5-year survival rate of just 7% in the United States [24]. To address the problem of automatic diagnosis of pancreatic cancer, we propose a new CAD framework for Intraductal Papillary Mucinous Neoplasms (IPMN). IPMN is a mucin-producing neoplasm found in the main and branch pancreatic ducts. They are radiographically identifiable precursors to pancreatic cancer [25, 26]. If left untreated, they can progress into invasive cancer. For instance, around one-third of resected IPMNs are found to be associated with invasive carcinoma [27]. In 2012, Tanaka et al. [28] published the International consensus guidelines for the preoperative management of IPMN using radiographic and clinical criteria. These guidelines can be used in the development of CAD approaches for the separation of IPMNs from normal pancreas. The CAD approaches can yield to identify important imaging biomarkers that may assist radiologists for improved diagnosis, staging, and treatment planning.

Our Contributions:

- To the best of our knowledge, this is the first study to use deep learning for the classification of IPMN.
- We employ multiple imaging modalities of MRI (T1 and T2) and fuse the feature representation using Canonical Correlation Analysis (CCA) to obtain better discrimination between normal and

subjects with IPMN. We also perform further stratification of IPMN in low-grade and high-grade categories, a necessary task for the management of IPMN.

- Extensive experimental evaluations are performed on a dataset comprising 139 subjects, the largest study of IPMN to date.

1.5 Unsupervised Learning Approaches for Tumor Characterization

Medical imaging researchers are continuously working on developing tumor characterization algorithms and are assisted by the availability of large publicly available datasets such as LIDC-IDRI from Lung Image Database Consortium [23] for evaluation. The dataset comprising 1018 with annotations from four radiologists is one of the largest lung nodule dataset. Hence, the developments and evaluations of supervised machine learning algorithms for lung nodule detection and classification are possible. Yet, at the same time, generating annotation labels for lung nodules remains a time-consuming task which is subject to intra and inter-observer variability among experts (radiologists).

In the final part of this dissertation, we explore the potential of unsupervised learning for tumor classification. Inspired by the successful application of unsupervised methods in other domains [29, 30], we study to extract important discriminative information from a large amount of unlabeled data. We employ both handcrafted and deep learning features extracted from pre-trained networks for classification. In order to obtain an initial set of labels in an unsupervised fashion, we cluster the samples into different groups in feature domain. We next propose to train *Proportion-Support Vector Machine ($\propto$SVM)* classifier using label proportions rather than instance labels. The trained model is then employed to classify testing samples as benign or malignant.

Our Contributions:

- To the best of our knowledge, our work is the first to study the potential of unsupervised learning methods for the lung nodule classification using both handcrafted and deep learning features.

- In order to alleviate the effect of noisy labels obtained during clustering, we propose to employ $\propto$SVM which is trained on label proportions. Instead of hard assigning label proportions, we estimate the label proportions in a data-driven manner. We experimentally establish that learning a discriminative model using the proposed combination of clustering and $\propto$SVM is superior in performance as compared to training a classifier with noisy instance level labels.

1.6 Dissertation Organization

This dissertation is organized as follows. In Chapter 2, a literature survey of different existing techniques regarding white adipose tissue, brown adipose tissue, supervised and unsupervised diagnosis techniques for lung nodules and IPMN is provided. In chapter 3, we present our segmentation and quantification approach for white adipose tissues. The detection and segmentation of brown fat are described in chapter 4. In chapter 5, we propose different supervised learning techniques for the malignancy determination of lung nodules and IPMN. Chapter 6 introduces unsupervised learning approaches for tumor characterization. Finally, in chapter 7, we present conclusions and future directions.

The dissertation comprises work on white and brown fat quantification published in [31, 32], as well as lung nodule and IPMN diagnosis published in [33–35]

CHAPTER 2: LITERATURE REVIEW

2.1 Fat Segmentation and Quantification

For abdominal fat (central obesity) quantification, Zhao et al. [36] used intensity profile along the radii connecting sparse points on the outer wall (skin boundary) starting from the abdominal body center. Boundary contour was then refined by a smoothness constraint to separate VAT from SAT. This method, however, does not adapt to obese patients easily, where the neighboring subcutaneous and visceral fat cavities lead to a leakage in the segmentation. In another study, Romero et al. [37] used heuristic search strategies to generate the abdominal wall mask on a small set of representative slices. In a similar fashion, the method in [38] developed a semi-supervised segmentation method based on a hierarchical fuzzy affinity function. Its success is vague when patient-specific quantification is considered. Mensink et al. [39] proposed to use a series of morphological operations; however fine-tuning of the algorithm is difficult, and the process should be repeated almost for every patient when the abdominal wall is too thin. More recently, Kim et al. [40] generated a subcutaneous fat mask using a modified "AND" operation on four different directed masks. However, logical and morphological operations make the whole quantification system vulnerable to inefficiencies. In a more advanced method such as in [41], SAT, VAT, and muscle were separated using a joint shape and appearance model, but the reproducibility of the method is highly dependent on the model at hand. Based on a similar idea as in [36], a recent method by Kim et al. [42] estimated the muscle boundary using a convex-hull and then performed smoothing by selecting points that minimize the distance between the contour and the organ regions. However, the performance is dependent on the goodness of fit of the convex-hull. Although the method addresses SAT-VAT separation at a volumetric level, it lacks the use of appearance features and volumetric smoothing which is important for improved segmentation.

Presently, there is no automated CAD system available for BAT quantification using radiology scans. Existing studies are mostly based on the qualitative observations of expert radiologists and nuclear medicine physicians. In those studies, strictly chosen specific anatomical locations were explored for BAT presence [43, 44]. The quantification process was conducted either by manual or semi-automated delineation methods. Since PET images have high contrast, thresholding and clustering-based methods are well-suited for the delineation of uptake regions. Therefore, a simple thresholding was often used for segmenting uptake regions pertaining to BAT, allowing the extraction of volumetric and SUV (i.e., "standardized uptake value") based metrics. BAT is considered present if there are areas of tissues that are (i) more than 5 mm in diameter, (ii) CT density is restricted to -190 to -30 Hounsfield Units (HU), and (iii) have an SUV of 18F-fluorodeoxyglucose (^{18}F-FDG) of at least 2 g/ml [43, 44] in corresponding PET images. Here it is important to note that in [45], authors chose the thresholding value for SUV_{max} >3 g/ml to identify BAT regions. Hence, there is no clear consensus on the choice of SUV for BAT regions. In the last step, regions of interest (ROIs) are manually defined to remove false positive (FP) regions from consideration. Several manual FP removal steps may be required for differentiating uptake between BAT regions and lymph nodes, vessels, bones, and the thyroid [46]. All these manual identifications require extensive user knowledge of the anatomy. Furthermore, in cases where pathologies are present, segregating pathologies from normal variants of ^{18}F-FDG on BAT regions can be extremely challenging [19].

2.2 Malignancy Characterization of Lung Nodules

Conventionally, the classification of lung nodules was performed using hand-crafted imaging features such as histograms [47], Scale Invariant Feature Transform (SIFT) [48], Local Binary Pat-

terns (LBP) [49] and Histogram of Oriented Gradients (HOG) [50]. The extracted sets of features were then classified using a variety of classifiers such as Support Vector Machines (SVM) [51] and Random Forests (RF) [52]. Recently with the success of deep convolutional neural network (CNN) for image classification, the detection and classification applications in medical imaging have adopted it for improved feature learning and representation. Tedious feature extraction and selection can now be circumvented using supervised high-level feature learning. This has also attracted the attention of researchers working in lung nodule detection and classification with limited success since the feature learning and classification were considered as separate modules. In those frameworks, a *pre-trained* CNN was only used for feature extraction whereas classification was based on an off-the-shelf classifier such as RF [53, 54]. In sharp contrast to these methods, we perform an end-to-end training of CNN for nodule characterization while combining multi-view features to obtain improved characterization performance.

Conventionally, the characterization of lung nodules comprised nodule segmentation, extraction of hand-crafted imaging features, followed by the application of an off-the-shelf classifier/regressor. The method by Uchiyama et al. [47] was based on the extraction of various physical measures, including intensity statistics and then classification using Artificial Neural Networks. El-Baz et al. [55] first segmented the lung nodules using appearance-based models and used spherical harmonic analysis to perform shape analysis. The final step was the classification using k-nearest neighbor. Proposing a study based on texture analysis, Han et al. [56] extracted 2D texture features such as Haralick, Gabor and Local Binary Patterns (LBP) and extended them to 3D. Support Vector Machine (SVM) was employed to perform the classification. In another classical work by Way et al. [57], segmentation is performed using 3D active contours followed by the extraction of texture features from the rubber band straightening transform of the surrounding voxels. The classification was performed using Linear Discriminant Analysis (LDA) classifier. In another study, Lee et al. [58] proposed a feature selection based approach using both imaging and clinical data. An

ensemble classifier, combining genetic algorithm (GA) and random subspace method (RSM) was then used to gauge feature relevance and information content. Finally, LDA was employed to perform classification on the reduced feature set.

Following up on the success of deep learning, the medical imaging community has moved from feature engineering to feature learning. In those frameworks, CNN had been used for feature extraction and an off-the-shelf classifier such as Random Forest (RF) was employed for classification [53, 54]. Recently, Buty et al. [54] combined spherical harmonics along with deep CNN features and then classified them using RF. However, the use of CNN for lung nodule classification has been confined to 2D image analysis [59], thus falling short of utilizing the important volumetric and contextual information.

Moreover, the use of high-level image attributes had been found to be instrumental in the risk assessment and classification of lung nodules. In an effort to study the relationship between nodules attributes and malignancy, Furuya et al. [60] found that in a particular dataset, 82% of the lobulated, 97% of the densely spiculated, 93% of the ragged and 100% of the halo nodules were malignant. Moreover, 66% of the round nodules were found to be benign. Inspired by this study, in this dissertation we utilize 3D CNN to a learn discriminative feature set corresponding to each of the 6 attributes. We then fuse these feature representations via MTL to determine the malignancy likelihood.

2.3 Diagnosis of IPMN

In the literature, there are a limited number of studies addressing the automatic diagnosis of IPMN using radiology images. Hanania et al. [61] studied the contribution of numerous low-level imaging features such as texture, intensity, and shape to perform low and high-grade IPMN classification.

In the approach by Gazit et al. [62] texture and component enhancing features were extracted from the segmented cysts. The process is then followed by a feature selection and classification framework. Both of these approaches [61, 62], however, are evaluated on CT images and require the segmentation of cysts or pancreas. In contrast to these methods, our approach doesn't require prior segmentation of cysts or pancreas and is evaluated on multi-modal MRI scans rather than CT. In this dissertation, we hypothesize and evaluate the influence of complementary information in T1-weighted and T2-weighted scans that can be utilized to perform improved diagnosis of IPMN.

2.4 Unsupervised Learning Approaches

Conventionally, the visual recognition and classification tasks were addressed using supervised learning algorithms which use labeled data. However, for tasks where manually generating labels corresponding to large datasets is laborious and expensive, the use of unsupervised learning methods is of significant value. Unsupervised techniques had been used to solve problems in various domains ranging from object categorization [29], narrated video summarization [30], text mining [63], and speech processing [64]. These methods typically relied on some complementary information provided with the data to improve learning which may not be available for many classification tasks in medical imaging.

In medical imaging, there have been different approaches that used unsupervised learning for detection and diagnosis problems. The approach by Shin et al. [65] used stacked autoencoders for multiple organ detection in MRI scans. Vaidhya et al. [66] presented a brain tumor segmentation method with stacked denoising autoencoder evaluated on multi-sequence MR images. In the work by Sivakumar et al. [67], the segmentation of lung nodules was performed with unsupervised clustering methods. In order to explore the information from unlabeled images, Zhang et al. [68]

described a semi-supervised method for the classification of four types of nodules. In sharp contrast to these approaches, the unsupervised learning strategies presented in this dissertation don't involve feature learning using autoencoders. Using the set of hand-crafted features, we propose an unsupervised learning approach where initially estimated labels are progressively improved. We also present a new unsupervised learning strategy with Generative Adversarial Networks (GANs).

CHAPTER 3: SEGMENTATION AND QUANTIFICATION OF WHITE ADIPOSE TISSUE

Related Publication:

Automatic Segmentation and Quantification of White and Brown Adipose Tissues from PET/CT Scans. Sarfaraz Hussein, A. Green, A. Watane, D. Reiter, X. Chen, G.Z. Papadakis, B. Wood, A. Cypess, M. Osman, U. Bagci. IEEE Transactions on Medical Imaging (IEEE-TMI) 36.3 (2017): 734-744.

This chapter is the first part of our exploration into adiposity-cancer association. In this chapter, we present a method for the segmentation and quantification of white adipose tissue. The segmentation and quantification framework for white adipose tissue begins with the detection of regions of interest. In this case, white fat segmentation is performed in the abdominal region. We then estimate the initial boundary that separates subcutaneous fat from the visceral fat. The boundary is iteratively improved using geometric median absolute deviation and appearance local outlier scores. The final step involves sparse 3D conditional random fields to obtain the final SAT-VAT separation. The following sections discuss these individual steps in detail.

3.1 Region Detection in whole-body CT volumes

The input to our region detection algorithm is a whole body CT volume $\mathcal{I} \in \mathbb{R}^{X \times Y \times Z}$. Since it is difficult to get a large amount of annotated data for training in medical imaging applications, one should resort to as few training examples as possible. Therefore, we propose a new region detection

method based on the concept of *one-shot learning*, as the learners are trained only on one image to make predictions for the remaining images. The proposed region detection framework locates two slices in the CT volume, i.e., top and bottom of the region of interest (yellow box in Figure 3.1). Detecting these two slices is challenging since they can easily be confused with similarly appearing slices. Therefore, there is a need for a better feature representation. In this regard, deep learning has recently adapted quite successfully for computer vision and medical imaging applications [69, 70]. To benefit from this rich representation of image features, we use Convolutional Neural Network (CNN) features (i.e., deep learning features) as image attributes extracted from the first fully-connected layer of the pre-trained Fast-VGG Network [71]. The network comprises 5 convolution layers and 3 fully connected layers. The first, second, and fifth convolution layers are followed by a max-pooling layer by convention. In order to have faster operations, 4-pixels stride is used in the first convolution layer. The dimension of the feature vector generated for each slice is equal to 4096. Given the reference annotations of the body regions for one subject's volumetric image, we find its Euclidean distance with the testing subjects' images using deep learning features. For training, we use two sets of learners: positive (D_p) and negative (D_n). The testing slice $I \in \mathcal{I}$ corresponding to the smallest distance with the positive set and largest distance with the negative set is selected as the desired result. In order to combine the probabilities pertaining to D_p and D_n learners, we use *logarithmic opinion pooling* [72] as:

$$P(I) = \frac{1}{Z} P(I|D_p)^w P(I|D_n)^{1-w}, \tag{3.1}$$

where $Z = \sum_{I \in \mathcal{I}} P(I|D_p)^w P(I|D_n)^{1-w}$ is the normalizing constant and w is the weight parameter.

3.2 SAT-VAT Separation and Quantification

The proposed SAT-VAT separation framework consists of 4 steps as illustrated in Figure 3.1. Since the HU interval for certain substances such as fat, water, and air in CT remains relatively constant, it is straightforward to identify TAT using a clinically validated thresholding interval on the HU space (step 1).

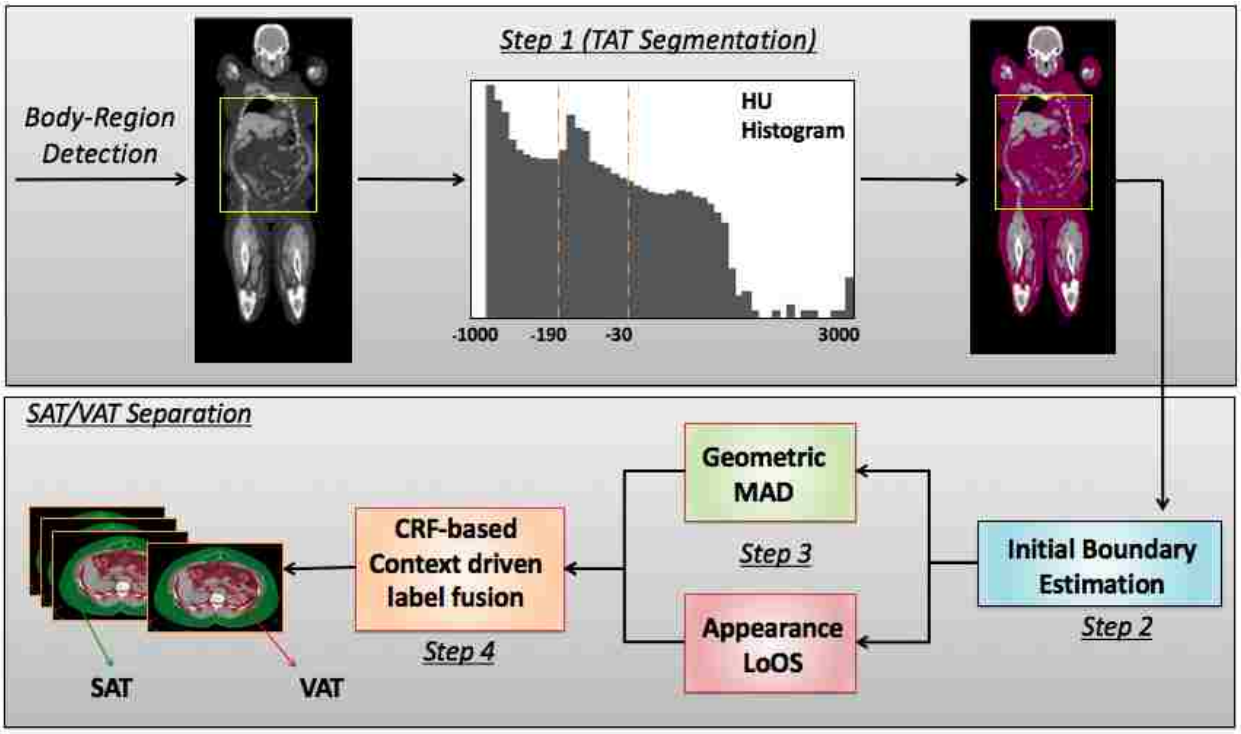


Figure 3.1: An overview of the proposed SAT-VAT separation method. Once the abdominal region is detected, Total Adipose Tissue (TAT) is segmented using the CT intensity interval known for fat tissue. Initial Subcutaneous-Visceral adipose tissue boundary is estimated by evaluating multiple hypothesis points. Geometric Median Absolute Deviation (MAD) and appearance based Local Outlier Scores (LoOS) are then combined within the 3D Conditional Random Field (CRF) based label fusion.

In step 2, we identify the initial boundary between VAT and SAT regions by conducting a sparse search over a line connecting the abdominal region center with the skin boundary (white dotted line in Figure 3.2). For step 3, two refinement methods are presented to remove FP boundary contour

points: Median Absolute Deviation (MAD) coefficient and Local Outlier Scores (LoOS). In the final step, we develop a sparse 3D CRF formulation to perform the finest SAT-VAT separation utilizing shape, anatomy, and appearance cues.

Step 1: Total Adipose Tissue (TAT) Segmentation

The input to our fat quantification pipeline is the abdominal volume. By following the clinical convention, we threshold the abdominal CT volume by -190 to -30 HU interval to obtain TAT [73]. We also perform a morphological closing on the input image using a disk with a fixed radius of r followed by a median filtering in an m x m neighborhood. This pre-processing is conducted to perform noise suppression and make the volume smooth for the next phase.

Step 2: Initial Boundary Estimation

We roughly identify the skin boundary of the abdominal region by selecting the longest isoline in the thresholded image (obtained from Step 1). For each point on the skin boundary contour $S = \{s_1, \ldots, s_n\}$, we generate a set of *hypotheses* $H = \{h_1, \ldots, h_u\}$ along the radii connecting S with its centroid C (Figure 3.2). Each hypothesis (candidate boundary location) is next verified for the possibility of being a boundary location by assessing image gradient information on the line connecting its location to the centroid C (white arrows in Figure 3.2). The SAT-VAT separation boundary, $B = \{b_1, \ldots, b_n\}$, should satisfy the following condition: $h_j \neq h_{j-1}$ for $h_j \in B$, and $b_i \in H$, $\forall i$. As illustrated in Figure 3.2, hypothesis points change their gradients in close vicinity of boundary B. These boundary points can still be noisy and may get stuck inside the small cavities of the subcutaneous fat. To alleviate such instabilities, the next step proposes a two-stage refinement methodology.

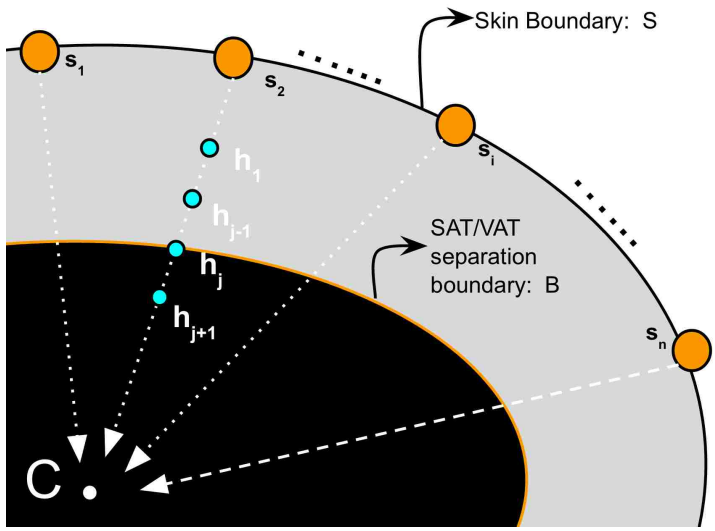


Figure 3.2: An illustration of skin boundary and hypothesis points along the radii connecting S with its centroid C. For each point (yellow) in S, a set of hypotheses (blue) is generated which is along the line connecting the skin boundary point with the centroid C.

Step 3: Outlier Rejection

Geometric MAD: In the first stage of the outlier removal, we apply MAD on the distances between B and S. The intuition behind this idea is that the SAT-VAT separation boundary should maintain a smoothly varying distance from the skin boundary. However, the outliers in subcutaneous and visceral cavities usually violate this smooth transition; therefore, we apply MAD on the points between B and S to remove outliers based on the geometric information. The higher outlier sensitivity of MAD in comparison with the mean-based method and other methods were studied in [74]. The resulting MAD coefficient Φ_i, for each boundary point, indicates a score for being an outlier:

$$\Phi_i = (|d_i - med(d)|)(med(|d_i - med(d)|))^{-1}, \tag{3.2}$$

where d is the Euclidean distance between S and B, $d_i = \|s_i - b_i\|_2$, and *med* is the median operator. Boundary locations with high MAD coefficients $\Phi > t$ (Section VI-A) are labeled as

outliers and subsequently removed from B.

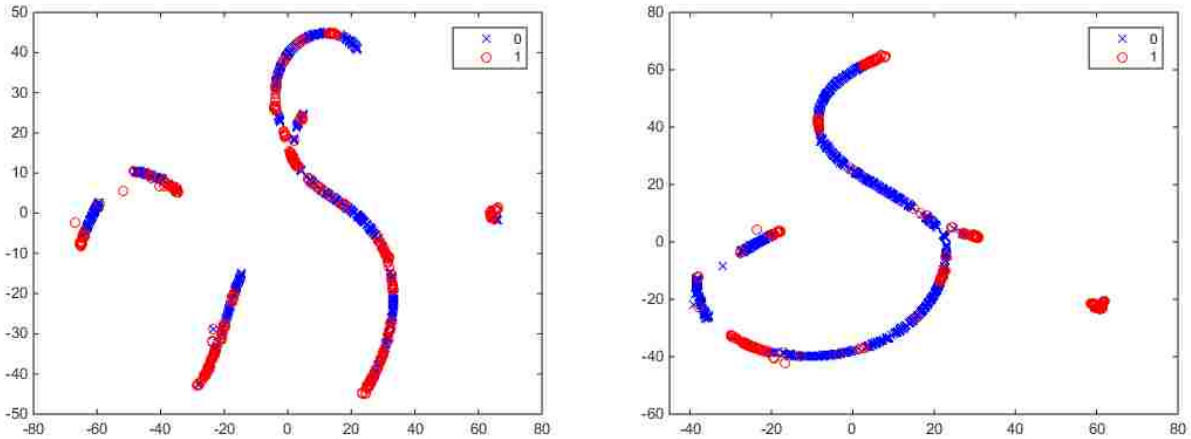
Local Outlier Scores: Although MAD can be quite effective in outlier rejection, there may still be some boundary locations that potentially lead to the drift of SAT-VAT separation due to the limitations of shape/geometry based attributes. To mitigate the influence of those boundary points, we apply the second stage of the outlier rejection, integrating appearance information through Histogram of Oriented Gradients (HOG) features [75]. For each candidate boundary point, we attach its appearance attributes (HOG) computed in a $c \times c$ cell. Since candidate boundary points lie on a high dimensional manifold (non-Euclidean), the *normalized correlation distance* computes similarities of those points. This is justified by computing the proximity, Q_{ij} between boundary points b_i and b_j using t-distributed stochastic neighborhood embedding (t-SNE) [76]:

$$Q_{ij} = \frac{1 + (\|b_i - b_j\|_2)^{-1}}{\sum_{u \neq v}(1 + (\|b_u - b_v\|_2)^{-1}}. \tag{3.3}$$

Figure 3.3 demonstrates the feature embedding visualization using t-SNE, where better separation of features with normalized correlation distance is observed. Points not mapped together to denser regions in high dimensional feature space are considered as outliers. By following this intuition, we obtain local outlier scores (LoOS) Π, indicating the confidence measure for each point being an outlier [77]:

$$\Pi(x) = erf\left(\frac{PLOF(x)}{\sqrt{2}.nPLOF}\right), \tag{3.4}$$

where erf is the Gaussian Error Function, and PLOF is the probabilistic local outlier factor based on the ratio of the density around point x and the mean value of estimated densities around all the remaining points. nPLOF is the λ standard deviation of the PLOF.



(a) **(b)**

Figure 3.3: t-SNE visualizations using (a) Euclidean and (b) Normalized Correlation distances. Better separation of classes (red and blue) can be clearly seen in (b).

Step 4: Context Driven Label Fusion using sparse 3D CRF

In order to fuse the labels of the boundary candidates across different slices of an image volume and create a fine SAT-VAT separating surface, we use sparse 3D Conditional Random Fields (CRF). In our CRF formulation, a set of N slices is selected to construct a graph $G = (V, E)$, where the nodes (V) consist of only the hypothesis boundary points (not the image pixels) and the edges (E) join neighboring boundary points in a high dimensional feature space. The labels, i.e., outlier and SAT-VAT boundary, are considered as source and sink in the context of our work. We define **unary potentials** of the CRF as the probabilities obtained after applying k-means clustering to the normalized outlier scores of the third stage:

$$\Theta(k_i|v_i) = -log(P(k_i|v_i)). \tag{3.5}$$

We define the **pairwise potentials** between the neighboring points v_i and v_j as:

$$\Psi(k_i, k_j|v_i, v_j) = \left(\frac{1}{1 + |\phi_i - \phi_j|}\right)[k_i \neq k_j], \tag{3.6}$$

where $|.|$ is the L_1 distance, $[.]$ is the indicator function, and ϕ is the concatenated vectorized appearance and geometric features. Once unary and pairwise potentials are defined, we seek to minimize the negative logarithm of $P(k|G; w)$ with k labels ($k \in \{0, 1\}$) and weights w as:

$$\begin{aligned} k^* &= \underset{k,w}{\operatorname{argmin}}(-log(P(k|G; w))) \\ &= \underset{k,w}{\operatorname{argmin}}(\sum_{v_i \in V} \Theta(k_i|v_i) + w \sum_{v_i, v_j \in E} \Psi(k_i, k_j|v_i, v_j)) \end{aligned} \tag{3.7}$$

Equation 3.7 is solved using graph-cut based energy minimization [78]. Graph-cut for more than two labels is an NP-hard problem and solved using approximate solutions. We have chosen graph-cut for minimizing the energy function defined to solve 3D sparse CRF. In contrast to level sets and loopy belief propagation methods, the graph-cut for two labels returns the global optimum in polynomial time. Additionally, graph cut formulation with a discrete binary solution space of [0,1] after linear programming relaxation (as in equation 3.7) is a convex problem. After solving equation 3.7, we fit a convex-hull around the obtained visceral boundaries and the segment inside the convex-hull is masked as VAT.

3.3 Results

3.3.1 Data for Central Obesity Quantification:

With IRB approval, we retrospectively collected imaging data from **151** subjects who underwent PET/CT scanning (67 men, 84 female, mean age: 57.4). Since CT images are from whole-body PET/CT scans (64-slice Gemini TF, Philips Medical Systems); they have low resolution, and no contrast agent was used for scanning. In-plane spacing (xy-plane) of CT image was recorded as 1.17 mm by 1.17 mm, and slice thickness was 5 mm. The scanner parameters for the CT were as follows: 120-140 kV and 33-100 mA (based on BMI), 0.5s per CT rotation, pitch of 0.9 and 512 $\times$ 512 data matrix was used for image fusion. The field of view (FOV) was from the top of the head to the bottom of the feet. The CT reconstruction process was based on filtered back-projection algorithm. No oral or intravenous contrast was administered.

Subjects were selected to have a *roughly equal* distribution of varying BMIs in order to have an unbiased evaluation. Our evaluation set comprised underweight subjects (N=20), normal subjects (N=50), overweight subjects (N=46), obese subjects (N= 35). UB ($>$ 10 years of experience in body imaging with CT and PET/CT interpretation) and GZP ($>$ 10 years of experience as a nuclear medicine physician and body imaging fellowship in radiology and imaging sciences) segmented fat regions by separating SAT and VAT boundary and using appropriate image post-processing such as edge-aware smoothing. Complementary to this interpretation, the participating radiologist BW ($>$ 20 years of experience in general radiology, body imaging, interventional radiology, and oncology imaging) evaluated SAT and VAT separating boundary qualitatively for both interpreters, and their segmentations were accepted at the clinical level of evaluations. This process is currently the most common procedure in creating a reference standard for segmentation evaluation [79–82].

Table 3.1: Abdominal region detection results measured by Intersection over Union (higher the better) and average absolute slice difference (lower the better) along with standard error of the mean (SEM). The IoUs are in 0-1 interval.

Methods	**IoU** *(SEM)*	**Avg.Abs. slice diff.** *(SEM)*
SIFT Flow [83]	0.263 *(0.019)*	90.22 *(2.71)*
Deep learning features [71] with Pos. learner	0.744 *(0.016)*	50.28 *(0.66)*
Proposed method (Equation 3.1)	**0.803** ***(0.014)***	**47.01** ***(0.62)***

Above 99% of agreement over Dice Similarity Coefficient (i.e. overlap ratio) was found between observers' evaluations with no statistical difference (t-test, $p > 0.5$).

3.3.2 Parameters and evaluations metrics:

The following parameters were noted for reproducible research in our experiments: $r = 10$, $m = \lambda = 3$, $t = 2.5$, $c = 14$, $w = 0.5$, and $N = 5$. For evaluation of region detection, we used *Intersection over Union (IoU)* [84] given by: $\frac{Overlap(R_G,R_S)}{max(|R_G|,|R_D|)}$, where R_G and R_D were reference standard and automatically detected abdominal region, respectively. We also reported region detection results using the absolute difference in slices between R_G and R_D. For segmentation evaluation, we used widely accepted Dice Similarity Coefficient (DSC): $\frac{2|I_G \cap I_S|}{|I_G|+|I_S|}$, where I_G and I_S were reference standard and automatically segmented fat region, respectively. Moreover, we measured the volumetric fat difference (in milliliters, mL) between true and segmented fat regions with Mean Absolute Error (MAE) metric.

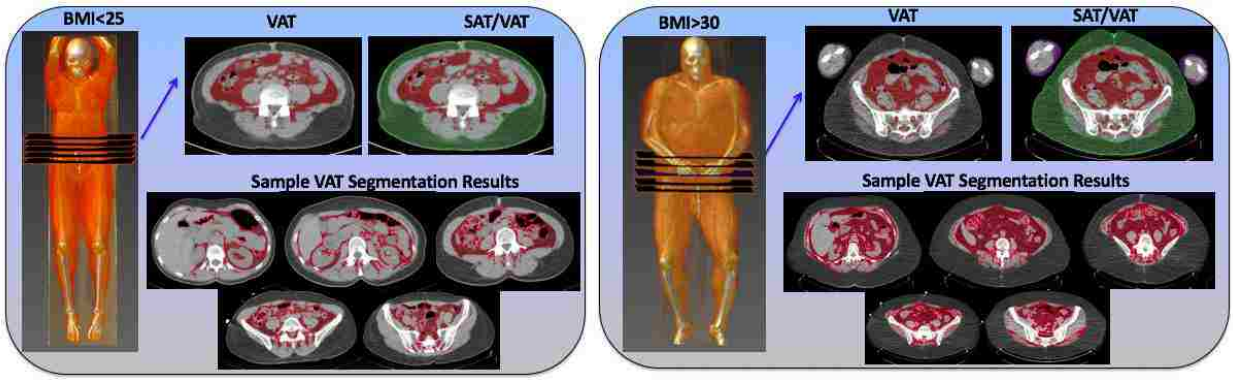


Figure 3.4: Visceral Adipose Tissue (red) and Subcutaneous Adipose Tissue (green) segmentations are illustrated for two subjects (one with BMI< 25, another with BMI> 30) at the chosen abdominal slice level along with their volume renderings. Several abdominal slices are also shown for central adiposity accumulation.

3.3.3 Comparisons:

For abdominal region detection, the upper boundary of the region was defined by the superior aspect of the liver, whereas the lower boundary was defined by the bifurcation of the abdominal aorta into the common iliac arteries [85]. As can be seen in Table 3.1, the proposed region detection method significantly outperformed registration based methods such as Scale Invariant Feature Transform (SIFT) flow [83]. Moreover, the proposed combination of positive and negative learners (Equation 3.1) provided percentage improvement of 7.9% in IoU and 6.5% reduction in average absolute slice difference, compared to only a positive learner with deep learning features.

We also performed extensive evaluations for SAT-VAT segmentation and quantification. We compared our method with One-class SVM, Zhao et al. [36], Random Sample Consensus (RANSAC) [86], and the state-of-the-art outlier detection method by Mahito et al. [87], which was based on iterative data sampling. In addition, we showed the results of our proposed framework's individual steps to provide a progressive improvement in accuracy, i.e., Geometric MAD, Appearance LoOS, and the final context driven fusion using sparse 3D CRF. As mentioned above, two delineations from expert interpreters were considered for the segmentation evaluation of SAT and VAT.

Table 3.2: Segmentation and quantification results for Subcutaneous Adipose Tissue (SAT) and Visceral Adipose Tissue (VAT) evaluated by Dice Similarity Coefficient (higher the better) and Mean Absolute Error (lower the better) along with standard error of the mean (SEM).

Methods	**SAT** *DSC (SEM)*	**VAT** *DSC (SEM)*	**SAT** *MAE in mL (SEM)*	**VAT** *MAE in mL (SEM)*
One-class SVM	0.886 *(0.004)*	0.842 *(0.006)*	16.695 *(0.963)*	16.696 *(0.963)*
Zhao et al. [36]	0.895 *(0.003)*	0.840 *(0.004)*	11.183 *(0.350)*	11.184 *(0.350)*
RANSAC [86]	0.913 *(0.003)*	0.861 *(0.005)*	14.126 *(1.179)*	14.126 *(1.179)*
Mahito et al. [87]	0.871 *(0.003)*	0.825 *(0.006)*	18.331 *(1.230)*	18.331 *(1.229)*
Geometric MAD	0.896 *(0.006)*	0.876 *(0.005)*	13.665 *(0.872)*	13.666 *(0.872)*
Appearance LoOS	0.925 *(0.002)*	0.885 *(0.004)*	11.815 *(0.857)*	11.816 *(0.856)*
Proposed method	**0.943** ***(0.003)***	**0.919** ***(0.003)***	**6.703** ***(0.466)***	**6.706** ***(0.466)***

Figure 3.4 shows the volume rendering of subjects along with VAT and SAT delineations for qualitative evaluation. Highly accumulated VAT (shown in red) in obese subjects is observed. DSC and MAE results for SAT and VAT are shown in Table 3.2 where significant improvements were obtained compared to other methods. The proposed method achieved around 40% lesser MAE, compared to [36] and other methods.

3.3.4 Computation Time:

The computation time for SAT-VAT segmentation method was less than $2s$/slice in our case, and less than $2.5s$/slice in other methods that we compared. The unoptimized MATLAB implementation of Geometric MAD took approximately $0.45s$/slice, that of appearance LoOS ran on average in $0.71s$/slice, followed by an average of $1.96s$/slice for 3D CRF on Intel Xeon Quad Core CPU @ 2.80GHz and 24.0GB RAM. Note also that none of the methods (in the comparison experiments)

required any manual intervention.

3.4 Summary

With obesity being one of the most prevalent health conditions in the world, its quantification, especially in the abdominal region is vital. In this regard, the quantification of visceral fat is significant. In parallel, since BAT is found to be negatively correlated with BMI [19], its quantification is essential for many clinical evaluations including obesity and metabolic syndromes. For central obesity quantification, we presented an unsupervised method for the separation of visceral and subcutaneous fat at the whole-body and body-region levels. In order to keep the proposed method fully automated, we also proposed a minimally supervised body region detection method where training was performed on a single subject. We ascribe the improved performance of our method to robust outlier rejection using geometric and appearance attributes followed by context driven label fusion. Evaluations were performed on non-contrast CT volumes from 151 subjects. Experimental results indicate that the proposed system has a great potential to aid in detecting and quantifying central obesity in routine clinical evaluations.

CHAPTER 4: BROWN FAT DETECTION AND SEGMENTATION

Related Publications:

- *Automatic Segmentation and Quantification of White and Brown Adipose Tissues from PET/CT Scans.* Sarfaraz Hussein, A. Green, A. Watane, D. Reiter, X. Chen, G.Z. Papadakis, B. Wood, A. Cypess, M. Osman, U. Bagci. IEEE Transactions on Medical Imaging (IEEE-TMI) 36.3 (2017): 734-744.

- *Brown adipose tissue detected by PET/CT imaging is associated with less central obesity.* A. Green, U. Bagci, Sarfaraz Hussein, P. Kelly, R. Muzaffar et al. Nuclear Medicine Communications (NMC) 38.7 (2017): 629-635.

Brown adipose tissue can be important to study the whole-body metabolism [19]. In this chapter, we describe an approach to detect and segment brown fat in whole-body PET-CT scans. In contrast to white fat, brown fat is metabolically active so it cannot be differentiated from white fat just with a CT image. The detection of brown fat, therefore, requires PET images as well. The region of interest for the detection of brown fat is confined to head/neck and thorax region. Our proposed BAT detection and delineation algorithm initiates with the segmentation of fat tissue from CT, followed by an automatic seed selection for BAT. We then perform PET-guided CT co-segmentation and lastly propose a false positive rejection method with novel probabilistic distance metrics. These 4 steps are depicted in Figure 4.1 and are explained in the next section.

4.1 Method for BAT Detection and Segmentation

Step 1: Segmenting Fat Tissue from CT Scans

The standard reference for estimating fat tissues in CT is by means of the computed planimetric method or with a fixed attenuation range from -190 to -30 HU (Chapter IV-Step 1). In our implementation, we extend this range to [-250, -10] HU in order to be more inclusive.

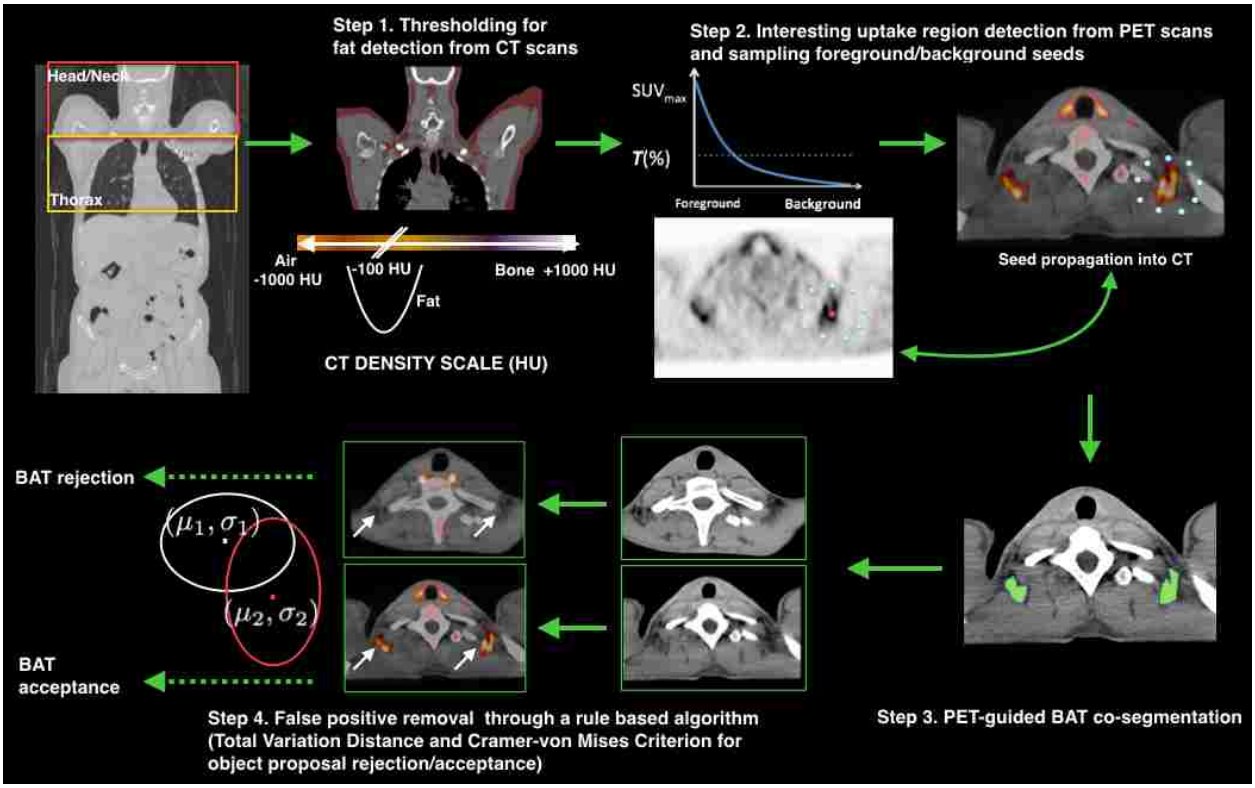


Figure 4.1: An overview of the proposed Brown Adipose Tissue (BAT) detection and segmentation system. Given the head-neck and thorax regions, adipose tissue is identified using CT thresholding intervals (Step 1). Using the corresponding PET scans, segmentation seeds are sampled in accordance with high uptake regions (Step 2). PET-CT co-segmentation is performed using Random Walk (Step 3) followed by false positive removal (Step 4) using Total Variation and Cramér-von Mises distances.

Prior to this operation, we employ a 3D median filtering to smooth the images. Resulting segmen-

tations will form the basis for differentiating BAT from non-BAT regions.

Step 2: Automatic Seed Selection

BAT regions are metabolically active, and studies reported that at least an $SUV_{max} = 2$ g/ml was observed in BAT regions [43, 44]. However, it is important to note that ^{18}F-FDG doesn't only attach to BAT but to tumor regions as well; hence, high SUV_{max} does not necessarily indicate BAT presence. To accurately characterize BAT, the anatomical/structural counterpart of the PET images is required. Since the BAT regions have $SUV_{max} \geq 2$ g/ml, we threshold the head/neck and thorax regions accordingly by following the automated seeding module of the joint segmentation method proposed in [88]. The resulting thresholded PET images most likely include numerous disconnected regions since many pixels may have SUV larger than 2 g/ml due to high metabolic activities. For each disconnected region, pixels with maximum SUVs are defined as foreground seeds. In order to set background pixels, we explore the neighborhood of foreground pixels by searching in 8-directions. We find background locations by marking the first pixel with less than or equal to 40% of the SUV_{max} (i.e., the conventional percentage for clinical PET thresholding). Those pixels are set as background seeds. The final step is to insert additional background seeds into the pixels lying in the spline connecting background seeds as explained in [88]. Once the background and foreground seeds are identified, Random Walk (RW) co-segmentation is employed by solving equation 4.2 for the unknown labels of the pixels (nodes).

Step 3: PET-guided Random Walk Image Co-Segmentation

It is reasonable to consider the BAT boundary determination process as a *co-segmentation problem* where the contributions of PET and CT in segmentation procedure are unequal. Based on our co-segmentation algorithm proposed in [88], herein we introduce a PET-guided RW co-segmentation

algorithm with asymmetric weights. This is based on the fact that the influence of PET on BAT segmentation results is higher than that of CT.

PET and CT images are in registration owing to PET/CT scanner's hybrid reconstruction properties. Any inconsistencies due to breathing and different timing of PET and CT imaging in the PET/CT scanner are minimized using deformable image registration as a post-processing step. Considering this fact, two graphs pertaining to CT and PET, $G^{CT} = (V^{CT}, E^{CT})$ and $G^{PET} = (V^{PET}, E^{PET})$, can be combined to define a hyper-graph $G^H = (V^H, E^H)$, on which the co-segmentation algorithm is applied. Note that for each image, we define a connected undirected graph G with nodes $v \in V$ and edges $e \in E \subseteq V \times V$. Since performing a random walk on the product graph G^H is equivalent to performing a simultaneous random walk on the graphs G^{CT} and G^{PET}, we define the nodes and edges as follows:

$$\begin{aligned} V^H =& \{(v_i^{CT}, v_i^{PET}) : v_i^{CT} \in V^{CT} \wedge v_i^{PET} \in V^{PET}\}, \\ E^H =& \{((v_i^{CT}, v_i^{PET}), (v_j^{CT}, v_j^{PET})) : \\ & (v_i^{CT}, v_j^{CT}) \in E^{CT} \wedge (v_i^{PET}, v_j^{PET}) \in E^{PET}\}. \end{aligned} \tag{4.1}$$

Similarly, the combinatorial Laplacian matrix definition L^H (that includes labeled and unlabeled nodes as well as weight parameters w of the imaging data) of the product graph G^H is updated from conventional RW formulation to co-segmentation as $L^H = (L^{CT})^\alpha \otimes (L^{PET})^\theta$, where α and θ are constants, $0 \leq \alpha < \theta \leq 1$, and $\otimes$ denotes direct product. Lastly, the probability distribution of intensity values for the product lattice x^H is defined as the direct multiplication of the initial probability distributions of x^{CT} and x^{PET} as $x^H = (x^{CT})^\zeta \otimes (x^{PET})^\eta$, where ζ and η are used to optimize the initial probability distributions subject to the constraint $0 \leq \zeta < \eta \leq 1$. The desired random walk probabilities are equivalent to the solution of the combinatorial Dirichlet

problem [89] as :

$$D[x^H] = \frac{1}{2}(x^H)^T L^H x^H = \frac{1}{2} \sum_{e_{ij} \in E^H} w_{ij}^H (x_i^H - x_j^H)^2, \tag{4.2}$$

where a combinatorial harmonic function of x^H, satisfying the Laplace equation $\bigtriangledown^2 x^H = 0$, minimizes equation 4.2. To emphasize the effect of PET more than that of CT for BAT region delineation, we select combination of (α, θ) as $(0.05, 0.95)$ and (ζ, η) as $(0.3, 0.7)$ after various empirical evaluations.

Step 4: Differentiating BAT from non-BAT regions

BAT regions are not easily separable from other fat regions in CT because WAT and BAT follow the same intensity patterns (fixed HU interval). Conventionally, intensity values of fat regions can be considered to follow a normal distribution with a known mean μ and variance σ (i.e., $\mathcal{C} = \mathcal{N}(\mu, \sigma)$). Since the PET and CT images are co-registered, the segmented regions in both PET and CT are equivalent: $r^{PET} = r^{CT}$. We next formulate the problem of differentiating BAT from non-BAT regions as follows. The intensity distribution p, obtained from r^{CT} correspondence of each segmented uptake region r^{PET}, should be in a close vicinity of $\mathcal{C}$, such that, $d(p, \mathcal{C}) < \epsilon$, where $d \in [0, 1]$ is a distance metric measuring whether p belongs to some class of distribution $\mathcal{C}$ or not. We postulate that p is sufficiently far from $\mathcal{C}$ when lymph nodes, tumor regions, or other non-fat tissues are involved in r^{CT}.

For the probabilistic distance metric in our framework (d), we use two complementary distance measures: total distance variation (d_{TV}) and Cramér-von Mises distance (d_{CM}). d_{TV} is equivalent

to the L_1-norm and can be formulated as half of the L_1-distance:

$$d_{TV} = \frac{1}{2} \sum_{x \in \Omega} |p(x) - \mathcal{C}(x)|, \tag{4.3}$$

where Ω is a measurable space on which p and $\mathcal{C}$ are defined. Complementary to d_{TV}, we also use d_{CM} to judge the *goodness of fit* between the two distributions by emphasizing L_2-distance. In other words, d_{CM} is effective in situations where two distributions under comparison have dissimilar shapes (although similar mean and variance can still be captured with d_{TV}). The Cramér-von Mises statistics is defined as:

$$d_{CM} = \underset{x}{min} |P(x) - \psi(x)|, \tag{4.4}$$

where $\psi(x)$ and $P(x)$ are cumulative distribution functions of $\mathcal{C}(x)$ and $p(x)$, respectively. The proposed distance d is simply formed by integrating d_{CM} and d_{TV} as:

$$d = \sqrt{d_{CM}^2 + d_{TV}^2}. \tag{4.5}$$

If $d < \epsilon$, our differentiation system accepts the BAT proposal/hypothesis. Note also that d is a *distance-metric* because (i) it is symmetric ($d(\mathcal{C}, p) = d(p, \mathcal{C})$), (ii) non-negative (as it spans from 0 to 1, $d \geq 0$), (iii) $d(p, \mathcal{C}) = 0$ only when $p = \mathcal{C}$, and (iv) it satisfies the triangle equality as:

$$d(p, \mathcal{C}) \leq d(p, \mathcal{D}) + d(\mathcal{D}, \mathcal{C}). \tag{4.6}$$

4.2 Results

4.2.1 Data for Quantification of BAT:

This retrospective study was institutional review board (IRB) approved and the need for written informed consent was waived. Thirty-seven adult (>21 years) oncology patients with FDG BAT uptake were identified from PET/CT studies from 2011-2013. The control cohort consisted of 74 adult oncology patients without detectable FDG BAT uptake matched for BMI/gender/season. The oncology patients have malignant tumors which were all biopsy proven. From the 4,458 FDG PET/CT reports in our database, there were 46 unique adult patients whose PET/CT reports specified the presence of BAT. Eight patients were excluded for only negligible PET/CT evidence of BAT reported in the paravertebral region. Another patient was excluded since FDG uptake was associated with interatrial lipomatous hypertrophy. Apart from these, the final selection of PET/CT scans was confirmed based on the consensus agreement of the participating nuclear medicine physicians, radiologist, and clinician. A total of 37 cases of adult BAT patients without FDG avid liver lesions were included in this study.

An intravenous injection of 5.18 MBq/kg (0.14 mCi/kg) ^{18}F-FDG was administered to patients with a blood glucose level $\leq$ 200 mg/dL after fasting for at least four hours. Patients sat in a quiet room during the 60 minute uptake phase and were instructed to remain quiet and refrain from physical activity. All scans were acquired using a Gemini TF (Philips Medical Systems) PET/CT scanner. There were no statistically significant differences between the two cohorts in gender, race, BMI, height, and weight ($p > 0.05$). The voxel dimensions in PET scans were 4 mm $\times$ 4 mm $\times$ 4 mm. The PET component of the PET/CT scanner was composed of lutetium-yttrium oxyorthosilicate (LYSO)-based crystal. Emission scans were acquired at 1-2 min per bed position. The FOV was from the top of the head to the bottom of the feet. The three-dimensional (3D)

whole-body acquisition parameters were 144×144 matrix and 18 cm FOV with a 50% overlap. The reconstruction process in the scanner was based on the 3D Row Action Maximum Likelihood Algorithm (RAMLA) [90].

To develop the reference standard, we used the manual delineation from three experts. First, the participating nuclear medicine physicians (MO: >20 years of experience, GZP: >10 years of experience, and AG: >10 years of experience), agreed on the predetermined SUV cut-off. GZP segmented the BAT regions, blind to consensus segmentation of MO and AG. Therefore, two delineations were considered in the evaluation, although three experts worked for the segmentation of BAT regions. When segmenting the BAT area, interpreters were provided viewer/fusion software, as well as manual, automated, and semi-automated contouring methods. The interpreters used both CT images (to define anatomical sites and fat tissue with the predefined HU interval) and PET images (with 2.0 g/ml of cut-off SUV_{max}) when delineating BAT regions. The fusion of PET with thresholded CT images provided uptake only in fat regions, removing most of the false positive uptake regions from consideration. Next, the interpreters used thresholding on PET uptake within an ROI (roughly drawn by the experts using manual contouring tool) for each detected BAT region. Finally, expert interpreters performed necessary corrections on the segmented PET uptake using manual contouring tools guaranteeing that the segmentations do not overlap with muscle, lymph nodes, and tumors.

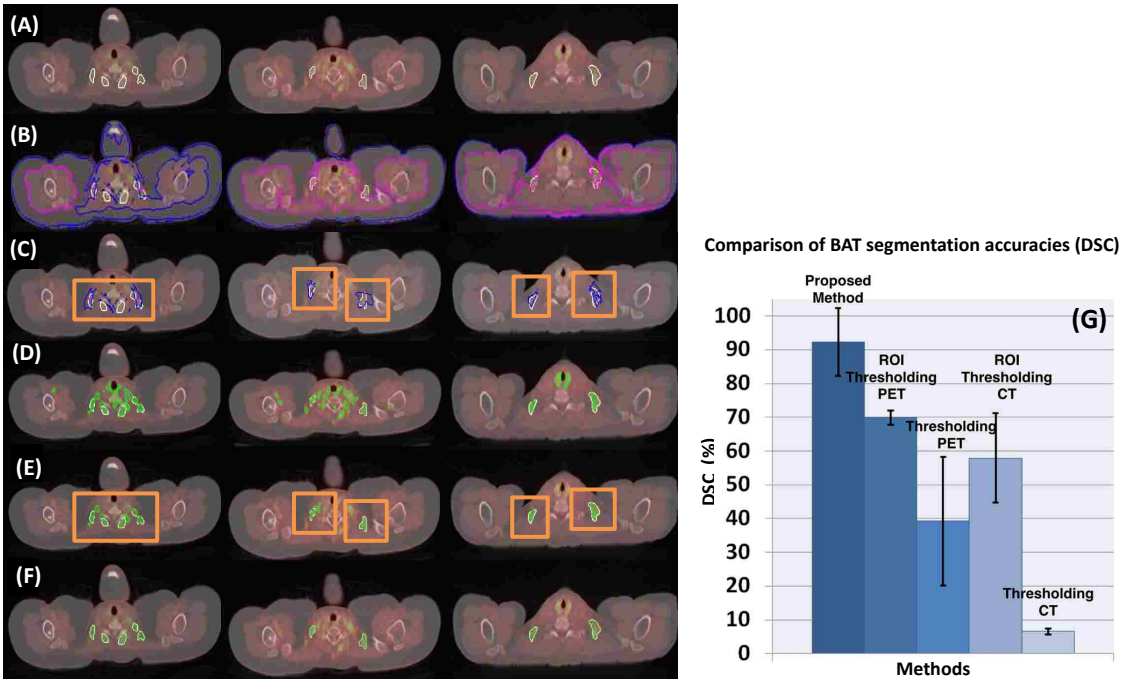


Figure 4.2: For three different anatomical levels (columns), row (A) shows reference standards (white); row (B) demonstrates the results from CT thresholding where pink (inner) and blue (outer) contours show brown fat delineation (blue contour shows fat region near skin boundary which leaks into the body cavity and also overlaps with pink contour as in the first column); row (C) comprises the results from ROI (Region of Interest) based CT thresholding, where orange boxes show user drawn ROIs and blue contours are the brown fat delineation results; row (D) shows the results from conventional PET thresholding, where green contours show output BAT delineations; row (E) depicts the ROI based PET thresholding; and row (F) demonstrates the proposed algorithm's delineation results using PET and CT jointly. (G) Dice Similarity Coefficients (DSC) of the proposed method in comparison with ROI based PET thresholding, PET thresholding, ROI based CT thresholding, and CT thresholding methods are shown.

4.2.2 Evaluation of Head-neck and Thorax Region Detection:

Anatomically, head-neck/thorax region was defined from the superior aspect of the globes to 5 mm below the base of the lungs [85]. We employed IoU as our region detection evaluation metric. Table 4.1 shows comparative evaluations of different methods with the proposed combination of positive and negative learners. The percentage improvement of 22.4% in IoU was observed over

SIFT Flow [83]. Moreover, the combination of positive and negative learners using logarithmic opinion pooling led to the percentage improvement of further 3% over the instance when only a positive learner was used.

4.2.3 Evaluation of BAT Delineation:

For quantitative evaluation of the delineation component of the proposed system, we compared True Positive (TP) and False Positive (FP) volume fractions of the segmented tissues with the manual delineation of the experts (mentioned in Section V).

We computed the average performance over the two delineations (Sensitivity (TP): 92.3 +/- 10.1%, Specificity (100-FP): 82.2+/-18.9 %). Metabolic volumes derived by the proposed segmentation were correlated with expert-derived volumes, and the resulting correlation coefficient, after linear regression, was found to be $R^2 = 0.97$ ($p < 0.01$). Example segmentation results at three different anatomical locations are shown in Figure 4.2A-F for qualitative evaluations. In the ROI based methods, ROIs were drawn by the user (expert annotator) to "roughly" include BAT regions, while excluding the rest (Figure 4.2-C and E).

4.2.4 Comparisons with Other Approaches:

We compared our method with the conventional thresholding approaches for both CT and PET (for SUV ≥ 2 g/ml) and then applied the logical AND operation to the two masks followed by a manual FP removal step. These were the methods used in previous studies to measure BAT boundary and volume [43]. Figure 4.2-G compares DSC of the proposed method with respect to the baseline

methods. Our proposed system outperformed other methods by a significant margin.

Table 4.1: Head-neck and Thorax Region detection results measured by Intersection over Union (IoU) and average absolute slice difference along with standard error of the mean (SEM). IoUs are in 0-1 interval.

Methods	**IoU** *(SEM)*	**Avg.Abs. slice diff.** *(SEM)*
SIFT Flow [83]	0.589 *(0.022)*	65.47 *(4.29)*
Deep learning features [71] with Positive learner	0.721 *(0.018)*	37.59 *(3.05)*
Proposed method (Equation 3.1)	**0.743** ***(0.006)***	**34.52** ***(1.28)***

4.2.5 Evaluation of BAT Region Proposals:

We computed TP and FP ratios over 111 PET/CT scans, each labeled as either BAT-positive or BAT-negative. Our results show that in 110 out of 111 scans (99.1%), BAT proposals' acceptance/rejection worked quite well. In only one scan, our system identified one region as non-BAT while the region was originally BAT. This false identification was due to significantly smaller size of the BAT region ($<$4mm), potentially due to the partial volume effect.

4.3 Summary

We offered a fully automated image analysis pipeline for BAT detection using PET/CT scans. Specifically, we proposed a novel approach to automatically detect and quantify BAT from PET/CT scans involving PET-guided CT co-segmentation, and a new probabilistic distance metric combining Total Variation and Cramér-von Mises distances. The proposed approach has a potential to assist in the clinical efforts to counteract obesity in the most natural way. We performed extensive

evaluations and our methods achieved state-of-the-art performances.

Since PET imaging provides biochemical and physiological activity, it remains the most accepted and preferred modality to study metabolically active BAT regardless of the radiation exposure. It is important to note that most of the BAT examples are obtained from the clinical trials or routine examination of different diseases. Moreover, there are a limited number of clinical trials solely focusing on BAT detection, quantification, and its role in metabolic syndrome, obesity, and other diseases. In order to reduce concerns regarding the ionizing radiation induced by PET/CT, one may consider reducing the radiation exposure of PET/CT scans. There are studies that show that low-dose CT scans have similar tissue HU levels as those in routine CT scans with no diagnostic differences noted, suggesting the use of low(er) dose CT scans in routine examinations [91]. On the other hand, lowering the radiation dose in PET equipment is more difficult and expensive than its CT counterpart [92,93]. Furthermore, the choice of a radiotracer is another concern while reducing the radiation dose. This is because the half-life of the most commonly used tracers is short and the patient size can affect image quality considerably [93]. Despite all the financial and logistical disadvantages, lowering the dose in the PET scans is a priority for the manufacturers, radiologists, and nuclear medicine physicians [92, 93]. With low dose PET/CT imaging, the cost-benefit ratio can be significantly improved for studies related to obesity and metabolic syndromes.

Other imaging modalities are also being explored for BAT detection and quantification. The application of MRI in human subjects is promising due to the lack of ionizing radiation and its excellent soft tissue contrast feature. However, current MR sequences do not have high sensitivity and specificity in identifying and quantifying BAT regions. Among a few works considering MR as a potential imaging modality for studying BAT, the use of Multi-point Dixon and multi-echo T2 spin MRI had been explored in mice [94]. Fuzzy c-means clustering was used for initial segmentation of BAT followed by a two-layer feed-forward neural network for the separation of BAT from WAT. However, high-field MRI is required for better separation of metabolically active fat regions from

the rest and there is no optimal sequence developed yet to do this task. Precise evaluation of BAT with MRI is not feasible in clinical routines and the current standards are still in favor of PET/CT.

Another alternative imaging modality to PET/CT for detection of BAT activation is contrast-enhanced ultrasound (CEUS) [95], a non-invasive and non-ionizing imaging modality. As the BAT activation was associated with an increased blood flow to the tissue, it can be measured by assessing the BAT perfusion. CEUS was found to detect increased BAT blood flow during cold exposure relative to warmer conditions. Although the reported experiments were preliminary with evaluations restricted to young and healthy males (mean age, 24.0 $\pm$ 2.4 years; mean body mass index, 23.4 $\pm$ 3.5 kg/m^2), BAT assessment may potentially be performed using CEUS in the future.

It should also be noted that the respiratory motion can be a potential source of error in co-segmentation. It is well known that the respiratory motion can affect PET and CT scans differently due to the possible differences in scan duration. This may induce residual registration mismatch between the two systems and eventually can lead to errors in BAT delineation. In such cases, motion correction algorithms as well as additional deformable registration methods can be employed to minimize registration errors prior to BAT segmentation.

Our study has some limitations to be noted. First, when young(er) subjects are scanned with their arms down, muscle may be observed as fat tissue due to photon depletion caused by high bone density. Although we did not observe this issue in the data set presented in this study, it may be a pressing issue that must be addressed when generalizing the quantification software into a larger cohort of studies such as clinical trials. Second, the partial volume effect can degrade the detection of small BAT deposits such as the para-spinal BAT, particularly when slice thickness in PET is large. Based on our recent findings in [96, 97], our future study will address these two limitations by integrating partial volume correction and denoising methods into the proposed system. Inspired

by a recent study [98], another step will be to design a fuzzy object modeling approach for the correction of incorrectly separated muscle and fat tissues due to photon depletion.

CHAPTER 5: SUPERVISED LEARNING FOR TUMOR CHARACTERIZATION

Related Publications:

- *TumorNET: Lung Nodule characterization using Multi-view Convolutional Neural Network with Gaussian Process.* Sarfaraz Hussein, Robert Gillies, Kunlin Cao, Qi Song, Ulas Bagci. IEEE International Symposium on Biomedical Imaging (ISBI 2017): pp. 1007-1010.

- *Risk Stratification of Lung Nodules using 3D CNN based Multi-task Learning.* Sarfaraz Hussein, Kunlin Cao, Qi Song and Ulas Bagci. Information Processing in Medical Imaging (IPMI 2017) pp. 249-260.

- *Deep Multi-Modal Classification of Intraductal Papillary Mucinous Neoplasms (IPMN) with Canonical Correlation Analysis.* Sarfaraz Hussein, Pujan Kandel, Juan E. Corral, Candice W. Bolan, Michael B. Wallace, Ulas Bagci. IEEE International Symposium on Biomedical Imaging (ISBI) 2018.

In this chapter, different supervised approaches for lung and IPMN are proposed. In the first part, we address the challenging problem of malignancy determination of lung nodules using 2D Convolutional Neural Network with Gaussian Process Regression. We also analyze the significance of high-level nodule attributes to improve malignancy determination. In the second part, we propose a 3D CNN network with graph-regularized multi-task learning (MTL) to improve risk stratification

of tumors. We also evade the need to have a large number of training examples by transfer learning. In the third part of this chapter, we propose a deep learning approach with canonical correlation analysis for the diagnosis of IPMN from multi-modal MR images.

5.1 TumorNET: Lung Nodule Characterization Using Multi-View 2D Convolutional Neural Network with Gaussian Process

We propose a new 2D Convolutional Neural Network (CNN) called TumorNET, where we train the network from scratch on lung nodules dataset. In order to generate sufficient number of training examples to train the network, we perform data-augmentation. Gaussian Process Regression (GPR) is employed in order to regress the malignancy score as an output.

5.1.1 Contributions:

- We perform an end-to-end training of CNN from scratch in order to realize the full potential of the neural network i.e. to learn discriminative features. Figure 5.1 shows an overview of the complete approach.
- Extensive experimental evaluations are performed on a dataset comprising lung nodules from more than 1000 low dose CT scans.
- We empirically establish the complementary significance of high-level nodule attributes such as calcification, lobulation, sphericity and others along with the deep CNN features to improve the malignancy determination.

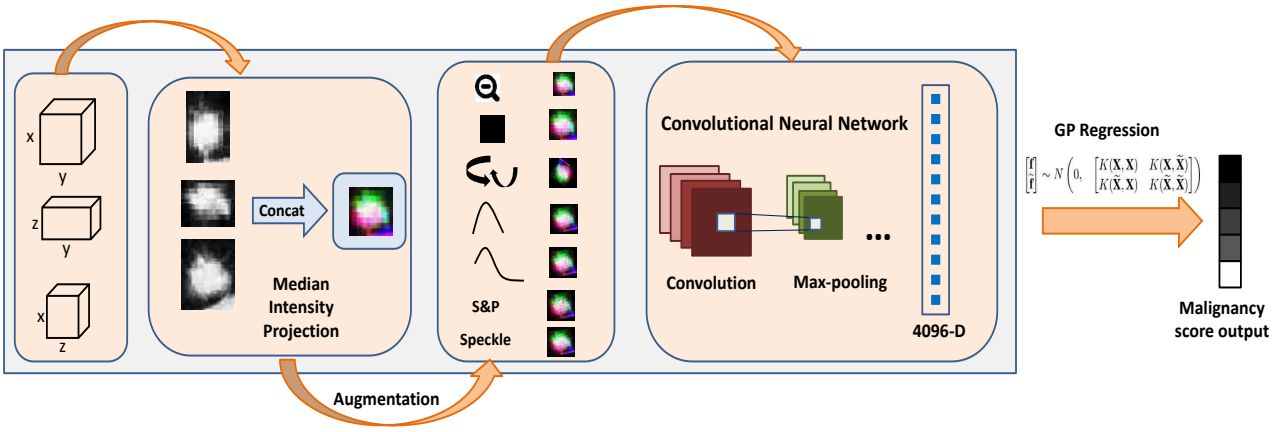


Figure 5.1: An overview of the proposed method. First, the median intensity projection is performed across each axis, followed by their concatenation as three channels of an image. Data augmentation is performed using scaling, rotation, adding Gaussian, Poisson, Salt and Pepper (S&P) and Speckle Noise. A CNN with 5 convolution and 3 fully connected layers is trained from scratch. For testing, the 3 channel image is passed through the trained network to get a 4096-dimensional feature from the first fully connected layer. Finally, the malignancy score is obtained using the Gaussian Process regression.

5.1.2 Materials

We evaluated our proposed approach on LIDC-IDRI dataset from Lung Image Database Consortium [23], which is one of the largest publicly available lung cancer screening dataset. The dataset comprises 1018 scans with slice thickness varying from 0.45 mm to 5.0 mm. Four expert radiologists annotated lung nodules with diameters greater than or equal to 3 mm. In our training and evaluation framework, we sampled nodules which were annotated by at least three radiologists. There were 1340 nodules satisfying this criterion. The nodules have a malignancy rating from 1 to 5 where 1 represents low malignancy and 5 is for highly malignant nodules. We excluded nodules with an average score equal to 3 to account for uncertainty in the decision of the radiologists. Our final dataset consists of 635 benign and 510 malignant nodules for classification. All images were resampled to have 0.5 mm spacing in each dimension.

5.1.3 Method

5.1.3.1 Multiview Deep Convolutional Neural Network:

Architecture:

Inspired by the success of deep convolutional neural networks for image classification we use a similar architecture as in [69] to perform end-to-end training of the CNN. *TumorNet* is the CNN network trained on the lung nodule dataset. The network is comprised of 5 convolutional layers, 3 fully connected layers and a softmax classification layer. The first, second, and fifth convolutional layers are followed by a max-pooling layer. Here it is important to note that CT image patches are in 3D, whereas the inputs to the network are 2D image patches. In order to combine information across all three views of the CT, we first compute the Median Intensity Projection of the image across each view. The median projected image ϕ corresponding to each dimension for an image patch I is given by:

$$\begin{aligned} \phi(y,z) &= \underset{x}{med}\,[I(x,y,z)], \\ \phi(x,z) &= \underset{y}{med}\,[I(x,y,z)], \\ \phi(x,y) &= \underset{z}{med}\,[I(x,y,z)], \end{aligned} \tag{5.1}$$

where *med* is the median operator. The 3 median projected images are then concatenated to form a 3-dimensional tensor $\Phi = [\phi(y,z), \phi(x,z), \phi(x,y)]$. This tensor Φ can be considered as a 2D image with 3 channels which is used to train the CNN network.

Data Augmentation:

Since the number of examples is insufficient to train a deep CNN network which often required a

large number of training examples, we generate extra training samples from the original data. The input data is augmented using rotation and scaling. We perform random rotation of the input patch along with two different scales (one for up-sampling and the other for down-sampling). Moreover, we also add Gaussian noise with random mean, Poisson, Salt & Pepper and Speckle noise. By applying this data augmentation strategy, we generate sufficient samples for both positive and negative examples to train our network.

5.1.3.2 Gaussian Process (GP) Regression:

The deep CNN network is trained using the augmented data until the loss converges for training data. In order to extract a high-level feature representation of the input data, we use the first fully connected layer of the network to obtain a $d = 4096$-dimensional feature vector. In order to reduce the computational cost, we randomly sample $n = 2000$ examples from the training data and extract their corresponding features. Let $\mathbf{X} = [X_1, X_2 \ldots X_n]$ be the feature matrix, where $\mathbf{X} \in \mathbb{R}^{n \times d}$. The regression labels are given by $\mathbf{Y} = [Y_1, Y_2 \ldots Y_n]$, where $\mathbf{Y} \in \mathbb{R}^{n \times 1}$. For label Y, we use the average malignancy scores from the radiologists which lie between 1 to 5, and the objective is to regress these scores in the testing data using the regression estimator learned from the training data.

As there exists inter-observer (radiologists) variation in the malignancy scores we model it with a Gaussian Process (GP), where the prediction for an input also comes with an uncertainty measure. In our GP formulation, each feature vector X_i is represented by a latent function f_i with $\mathbf{f} = (f_1, f_2 \ldots f_n)$ which is defined as:

$$\mathbf{f}|\mathbf{X} \sim N(m(\mathbf{X}), K(\mathbf{X},\mathbf{X})), \tag{5.2}$$

where $m(\mathbf{X})$ is the mean function and K is the covariance matrix such that $K_{ij} = k(X_i, X_j)$. The GP regression, corresponding to a single observation Y is modeled by a latent function and Gaussian noise ϵ:

$$Y = f(X) + \epsilon, \epsilon \sim N(0, \sigma_n^2). \tag{5.3}$$

If $\mathbf{f}$ and $\widetilde{\mathbf{f}}$ represent training and testing outputs, then their joint distribution is given by:

$$\begin{bmatrix} \mathbf{f} \\ \widetilde{\mathbf{f}} \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) & K(\mathbf{X}, \widetilde{\mathbf{X}}) \\ K(\widetilde{\mathbf{X}}, \mathbf{X}) & K(\widetilde{\mathbf{X}}, \widetilde{\mathbf{X}}) \end{bmatrix}\right), \tag{5.4}$$

where $K(\widetilde{\mathbf{X}}, \mathbf{X})$ represent the covariances evaluated between all pairs of training and testing sets. Finally, the best estimator for $\widetilde{\mathbf{f}}$ is computed from the mean of this distribution.

5.1.4 Experiments and Results

For evaluations, we performed 10 fold cross validation over 1145 nodules. The proposed data augmentation yielded 50 extra samples corresponding to each example in the training data. We used an equal number of positive and negative examples to perform balanced training of the network without any bias. From the training set, we sampled 10% examples to be used as validation for the CNN. The network was trained for approximately 10,000 iterations as the loss function converged around it.

After the network was trained, we randomly sampled 2000 examples from the training data and extracted features corresponding to the first fully connected layer of the network. The GP regression was then applied to those features. The images from the test set were forward passed through the network to obtain the same feature representation followed by GP regression.

Table 5.1: Comparison of the proposed approach with Support Vector Regression, Elastic Net and LASSO using accuracy measure and standard error of the mean (SEM).

Methods	**Regression Acc.% (SEM%)**
Support Vector Regression	79.91 (1.36)
Elastic Net	79.74 (0.94)
LASSO	79.56 (1.14)
GP Regression (Proposed)	**82.47 (0.62)**

A nodule was said to be classified successfully if its predicted score lies in ± 1 margin of the true score. This was done to account for any inter-observer variability in the dataset. Comparisons were performed using Support Vector Regression, Elastic Net and Least Absolute Shrinkage and Selection Operator (LASSO), where CNN features were used in all these methods. As it can be inferred from Table 5.1 that TumorNet with GP regression outperforms popular classification and regression methods by a significant margin. Sample qualitative results are visualized in Figure 5.2.

High level Nodule Attributes:

We also explored the significance of high-level nodule attributes such as calcification, sphericity, texture, and others for the determination of nodule malignancy. Fortunately, for the LIDC-IDRI dataset, the radiologists have also provided the scores corresponding to each of these attributes for nodules larger than 3 mm. We aim to analyze how these high-level attributes can aid classification of a nodule in conjunction with the appearance features obtained using the TumorNet framework. Another reason for our interest in these high-level attributes is that they can be easier to detect and annotate as compared to malignancy. In this regard, crowdsourcing can be employed to get these attributes with high efficiency and efficacy. For this particular experiment, we used 6 attributes, i.e., calcification, spiculation, lobulation, margin, sphericity, and texture. We computed the average scores in cases where scores from multiple radiologists were available.

Table 5.2: Regression accuracy and standard error (SEM) using the combination of high-level attributes and CNN features.

Methods	**Regression Acc.% (SEM%)**
High level attributes	86.58 (0.59)
High level attributes+CNN	**92.31 (1.59)**

We performed two sets of experiments. For first we used GP regression over the set of these 6 features and for second we concatenated them with 4096 dimension feature vector from TumorNet. We found that the combination of the high-level attributes and CNN features notably improves the regression accuracy (Table 5.2).

5.1.5 Summary

In contrast to the previous studies that used pre-trained network [53,54], in this work, we proposed an end-to-end training of deep multi-view Convolutional Neural Network for nodule malignancy determination termed TumorNet. In order to cater to the need to have a large amount of labeled data for CNN, we performed data augmentation using scale, rotation and different categories of noise. In order to combine 3 different views from the image, we performed median intensity projection followed by their concatenation in a tensor form of a single image with 3 channels.

Furthermore, we explored the significance of high-level nodule attributes for malignancy determination.

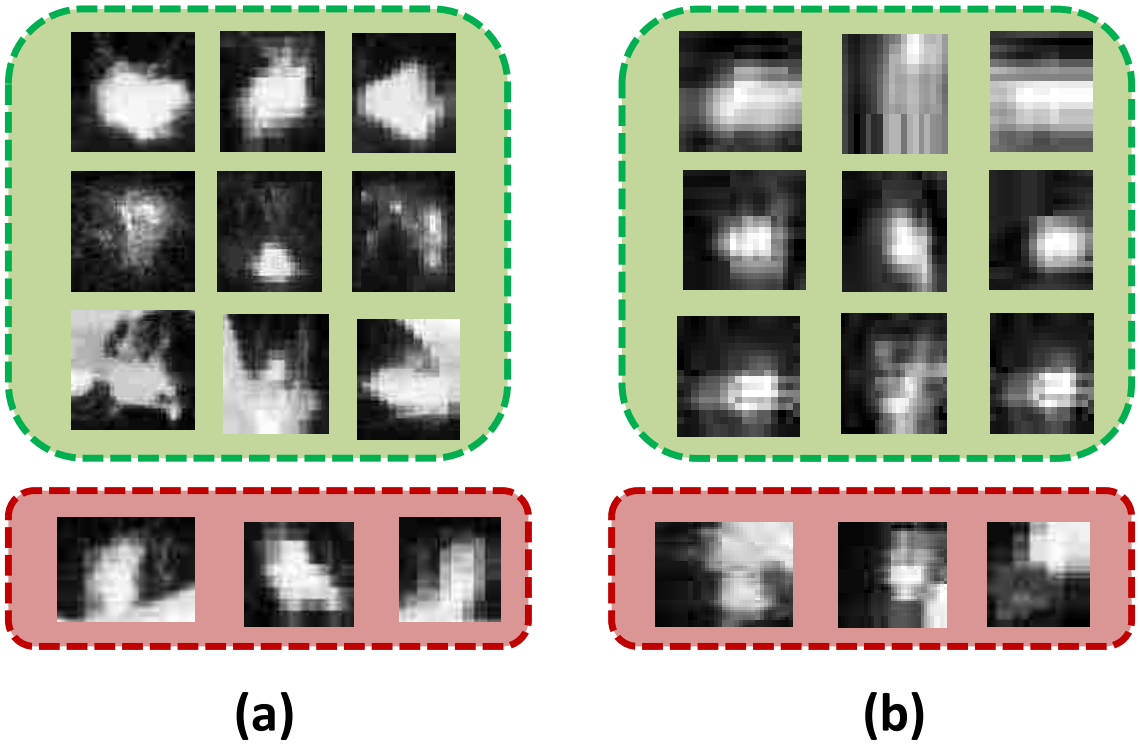


Figure 5.2: Qualitative results showing median intensity projected images for correctly (green) and incorrectly (red) scored lung nodules. (a) and (b) show malignant and benign nodules respectively where each row shows different cases and column represents different views (axial, sagittal, coronal).

We found that these attributes are of high significance and are actually complementary to the multi-view deep learning features. We obtained a substantial improvement in accuracy using the combination of both high-level attributes and CNN features.

As an extension to this study, our future work will involve the automatic detection of high-level nodule attributes and their use for malignancy determination. As these attributes may not be specific to radiology, transfer learning from other computer vision tasks can assist in addressing the challenge of the unavailability of a large amount of labeled data in radiology. Moreover, other

imaging modalities such as PET could be considered for diagnostic imaging of lung cancer and treatment planning within the TumorNet framework.

5.2 Malignancy characterization of Lung Nodules using 3D CNN Based Multi-Task Learning:

In this part of the dissertation, we address the challenge of risk-stratification of lung nodules in low-dose CT scans. Capitalizing on the significant progress of deep learning technologies for image classification and their potential applications in radiology [70], we propose a 3D Convolutional Neural Network (CNN) based approach for rich feature representation of lung nodules. We argue that the use of 3D CNN is paramount in the classification of lung nodules in low-dose CT scans which are 3D by nature. By using the conventional CNN methods, however, we implicitly lose the important volumetric information which can be very significant for accurate risk stratification. The superior performance of 3D CNN over 2D networks is well studied in [99]. We also avoid hand-crafted feature extraction, painstaking feature engineering, and parameter tuning. Moreover, any information about six high-level nodule attributes such as calcification, sphericity, margin, lobulation, spiculation, and texture (Figure 5.3) can help in improving the benign-malignant risk assessment of the nodules. Taking forward this idea, we identify features corresponding to these high-level nodule attributes and fuse them in a multi-task learning framework to obtain the final risk assessment scores. An overview of the proposed approach is presented in Figure 5.4. Overall, our main contributions in this work can be summarized as follows:

- We propose a 3D CNN based method to utilize the volumetric information from a CT scan which would be otherwise lost in the conventional 2D CNN based approaches. Moreover, we also circumvent the need for a large amount of volumetric training data to train the 3D network by transfer learning. We use the CT data to fine-tune a network which is trained on 1 million videos.

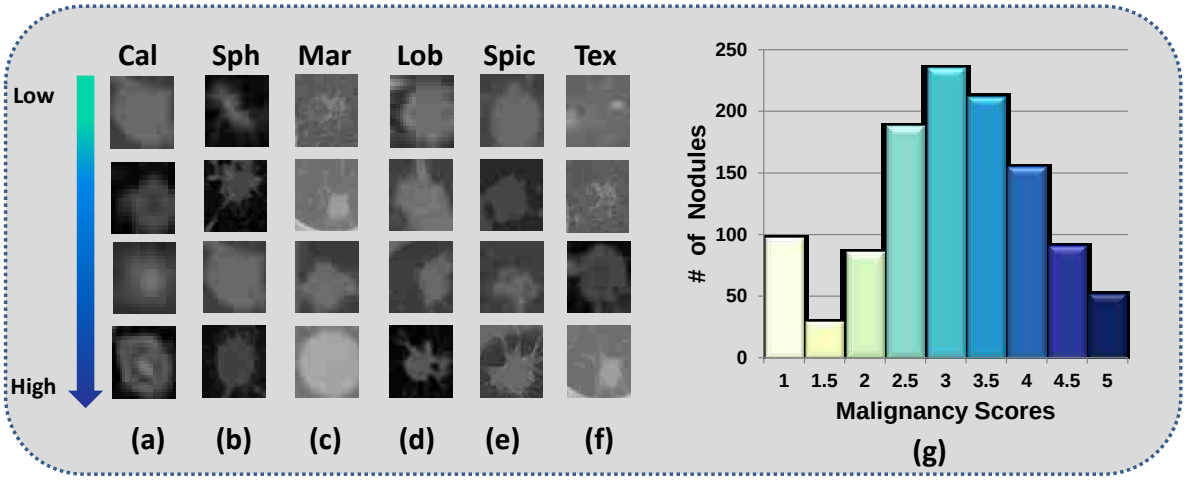


Figure 5.3: Lung nodule attributes with different scores. As we move from the top (attribute missing) to the bottom (attribute with the highest prominence), the prominence of the attributes increases. Column (a) and (b) show calcified and spherical nodules; (c) represents margin where the top row is for poorly defined nodules and the bottom row shows well-defined nodules. Column (d) and (e) show lobulated and spiculated nodules whereas (f) represents nodules with different textures. The top row in (f) represents a non-solid nodule and the bottom row shows a solid nodule. The graph in (g) shows the number of nodules with different malignancy scores

To the best of our knowledge, our work is the first to empirically validate the success of transfer learning of a 3D network for lung nodules.

- We perform experimental evaluations on one of the largest publicly available datasets comprising lung nodules from more than 1000 low-dose CT scans.

- We employ graph regularized sparse multi-task learning to fuse the complementary feature information from high-level nodule attributes for malignancy determination. We also propose a scoring function to measure the inconsistency in risk assessment among different experts (radiologists).

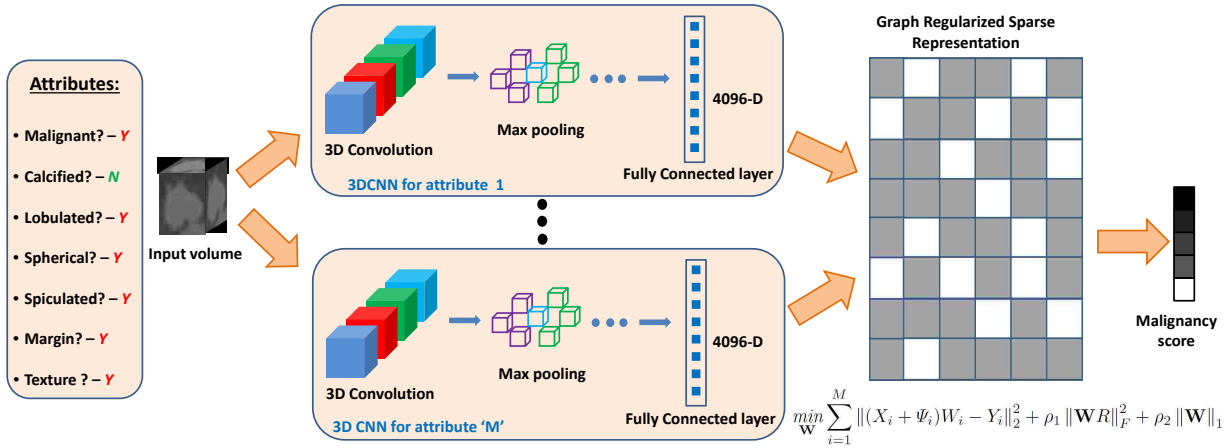


Figure 5.4: An overview of the proposed approach. First, we fine-tune 3D CNNs using labels for malignancy and six attributes. Given the input volume, we pass it through different 3D CNNs each corresponding to an attribute (task). The network comprises 5 convolution, 5 max pooling, and 2 fully connected layers. We use the output from the first fully connected layer as the feature representation. The features from different CNNs are fused together using graph regularized sparse least square optimization function to obtain coefficient vectors corresponding to each task. During the testing phase, we multiply the feature representation of the testing image with the coefficient vector to obtain the malignancy score.

5.2.1 Method

5.2.1.1 Problem Formulation

Let $X = [x_1, x_2 \dots x_n] \in \mathbb{R}^{n \times d}$ be the data matrix comprising features from n data points in $\mathbb{R}^d$. Each sample corresponds with a regression score given by $Y = [y_1, y_2 \dots y_n]$ where $Y \in \mathbb{R}^{n \times 1}$. Here the objective is to learn the coefficient vector or the regression estimator W from the training data. In this case, the ℓ_1 regularized least square regressor is defined as:

$$\min_{W} \|XW - Y\|_2^2 + \lambda \|W\|_1 , \tag{5.5}$$

where λ controls the sparsity level for coefficient vector $W = [w_1, w_2 \ldots w_d]$. The problem in Eq. 5.5 is an *unconstrained convex optimization* problem, and it remains non-differentiable when $w_i = 0$. Hence, the closed form solution corresponding to the global minimum for Eq. 5.5 is not possible. Thus, the above equation is represented in the following way as a *constrained optimization* function:

$$\begin{aligned} \min_{W} \|XW - Y\|_2^2 \,, \\ s.t. \, \|W\|_1 \leq t, \end{aligned} \tag{5.6}$$

where t is inversely proportional to λ. In the representation given in Eq. 5.6, both optimization function and the constraint are convex.

5.2.1.2 Network Architecture and Transfer Learning

We use the lung nodules dataset to fine-tune a 3D CNN trained on Sports-1M dataset [100]. The sports dataset comprises 1 million videos with 487 classes. In the absence of a large number of training examples from lung nodules, we use transfer learning strategy to obtain rich feature representation from a larger dataset (Sports-1M) for lung nodule characterization. The Sports-1M dataset is used to train a 3D CNN [99]. The network comprises 5 convolution, 5 max-pooling, 2 fully-connected and 1 soft-max classification layers. The input to the network is $3 \times 16 \times 128 \times 171$ where there are 16 non-overlapping slices in the input volume. The first 2 convolution layers have 64 and 128 filters respectively, whereas there are 256 filters in the last 3 layers. The outputs of the fully connected layers are of 4096 dimensions.

5.2.1.3 Multi-task learning

Consider a problem with M tasks representing different attributes corresponding to a given dataset D. These tasks may be related and share some feature representation, both of which are unknown. The goal in Multi-task learning (MTL) is to perform joint learning of these tasks while exploiting dependencies in feature space so as to improve regressing one task using the others. In contrast to multi-label learning, tasks may have different features in MTL. Each task has model parameters denoted by W_m, used to regress the corresponding task m. Moreover, when $\mathbf{W} = [W_1, W_2 \dots W_M] \in \mathbb{R}^{M \times d}$ represents a rectangular matrix, rank is considered as an extension to the cardinality. In that case, trace norm, which is the sum of singular values is a replacement to the ℓ_1-norm. Trace norm also known as the nuclear norm is the convex envelope of the rank of a matrix (which is non-convex), where the matrices are considered on a unit ball. By replacing ℓ_1-norm with trace norm in Eq. 5.5, the trace norm regularized least square loss function is given by:

$$\min_{\mathbf{W}} \sum_{i=1}^{M} \|X_i W_i - Y_i\|_2^2 + \rho \|\mathbf{W}\|_* , \tag{5.7}$$

where ρ tunes the rank of the matrix $\mathbf{W}$, and trace-norm is defined as: $\|\mathbf{W}\|_* = \sum_{i=1} \sigma_i(\mathbf{W})$ with σ representing singular values.

Another regularizer, pertinent to MTL, is the regularization on the graph representing the relationship between the tasks [101, 102]. Consider a complete graph $G = (V, \mathcal{E})$, such that nodes V represent the tasks and the edges $\mathcal{E}$ encode any relativity between the tasks. The complete graph can be represented as a structure matrix $S = [e^1, e^2 \dots e^{\|\mathcal{E}\|}]$ and the difference between all the

pairs connected in the graph is penalized by the following regularizer:

$$\|\mathbf{W}S\|_F^2 = \sum_{i=1}^{\|\mathcal{E}\|} \left\|\mathbf{W}e^i\right\|_2^2 = \sum_{i=1}^{\|\mathcal{E}\|} \left\|\mathbf{W}_{e_a^i} - \mathbf{W}_{e_b^i}\right\|_2^2. \tag{5.8}$$

Herein, e_a^i, e_b^i are the edges between the nodes a and b. The above regularizer can also be written as:

$$\|\mathbf{W}S\|_F^2 = \text{tr}((\mathbf{W}S)^T(\mathbf{W}S)) = \text{tr}(\mathbf{W}SS^T\mathbf{W}^T) = \text{tr}(\mathbf{W}\mathcal{L}\mathbf{W}^T), \tag{5.9}$$

where 'tr' represents trace of a matrix and $\mathcal{L} = SS^T$ is the Laplacian matrix. Since there may exist disagreements between the scores from different experts (radiologists), we propose a scoring function to measure potential inconsistencies:

$$\Psi(j) = (e^{\frac{-\sum_i (x_i^j - \mu^j)^2}{2\sigma^j}})^{-1}. \tag{5.10}$$

The inconsistency measure corresponding to a particular example j is represented by $\Psi(j)$. x_i^j is the score given by the expert (radiologist) i and μ^j and σ^j denote mean and standard deviation of the scores, respectively. Here, for simplicity, we have dropped the index for the task; however, note that the inconsistency measure is computed for all the tasks. The final proposed graph regularized sparse least square optimization function with the inconsistency measure can then be written as:

$$\min_{\mathbf{W}} \sum_{i=1}^{M} \overbrace{\|(X_i + \Psi_i)W_i - Y_i\|_2^2}^{①} + \overbrace{\rho_1 \|\mathbf{W}S\|_F^2}^{②} + \overbrace{\rho_2 \|\mathbf{W}\|_1}^{③}, \tag{5.11}$$

Algorithm 1 Algorithm for the proposed MTL method

Input: Generated features from 3D CNN: X_M^N for M attributes and N examples
Attributes scores: Y_M^N
Output: Coefficient matrix $\mathbf{W}$
Step 1 **– for** each task $i = 1$ to M and each example $j = 1$ to N **do**
Solve equation (5.10) to find Ψ
end for
Step 2 **–** Formulate objective function as in equation (5.11)
Step 3 **–** Use accelerated proximal gradient method to optimize equation (5.11)
return $\mathbf{W}$

where ρ_1 controls the level of penalty for graph structure and ρ_2 controls the sparsity. In the above optimization, the least square loss function ① considers tasks to be decoupled whereas ② and ③ consider the interdependencies between different tasks.

5.2.1.4 Optimization

The optimization function in Eq. 5.11 cannot be solved through standard gradient descent because the ℓ_1-norm is not differentiable at $\mathbf{W} = 0$. Since the optimization function in Eq. 5.11 has both smooth and non-smooth convex parts, estimating the non-smooth part can help solve the optimization function. Therefore, *accelerated proximal gradient method* [103, 104] is employed to solve the Eq. 5.11. The accelerated proximal method is the first order gradient method with a complexity of $O(1/k^2)$, where k is the iteration counter. Note that in Eq. 5.11, the ℓ_1-norm comprises the non-smooth part and the proximal operator is used for its estimation. The steps in the proposed approach are summarized in Algorithm 1.

5.2.2 Experiments

5.2.2.1 Data

For evaluating our proposed approach, we used LIDC-IDRI dataset from Lung Image Database Consortium [23], which is one of the largest publicly available lung cancer screening datasets. There were 1018 CT scans in the dataset, where the slice thickness varied from 0.45 mm to 5.0 mm. The nodules having diameters larger than or equal to 3 mm were annotated by at most four radiologists.

The nodules which were annotated by at least three radiologists were used for the evaluations. There were 1340 nodules satisfying this criterion. We used the mean malignancy and attribute scores of different radiologists for experiments. The nodules have ratings corresponding to malignancy and the other six attributes which are (i) calcification, (ii) lobulation, (iii) spiculation, (iv) sphericity, (v) margin and (vi) texture. The malignancy ratings varied from 1 to 5 where 1 indicated benign and 5 represented highly malignant nodules. We excluded nodules with an average score equal to 3 to account for the indecision among the radiologists. Our final dataset consisted of 635 benign and 509 malignant nodules for classification. The images were resampled to have 0.5 mm spacing in each dimension.

5.2.2.2 Results

We used the 3D CNN trained on Sports-1M dataset [100] which had 487 classes. We fine-tuned the network using samples from lung nodule dataset. In order to generate the binary labels for the six attributes and the malignancy, we used the center point and gave positive (or negative) labels to samples having scores greater (or lesser) than the center point. In the context of our work, tasks

represented six attributes and malignancy. We fine-tuned the network with these 7 tasks and performed 10 fold cross-validation. By fine-tuning the network, we circumvented the need to have a large amount of training data. Since the 3D network was trained on image sequences with 3 channels and with at least 16 frames, we replicated the same gray level axial channel for the other two.

Table 5.3: Classification accuracy and mean absolute score difference (lower the better) of the proposed multi-task learning method in comparison with the other methods.

Methods	**Accuracy**	**Mean score diff.**
GIST features with LASSO	76.83%	0.6753
3D CNN MTL with Trace norm	80.08%	0.6259
Proposed method (Equation 5.11)	**91.26%**	**0.4593**

Moreover, we also ensured that all input volumes have 16 slices by interpolation when necessary. We used the 4096-dimensional output from the first fully connected layer of the 3D CNN as a feature representation.

To find the structure matrix S, we computed the correlation between tasks by finding an initial normalized coefficient matrix $\mathbf{W}$ using lasso with least square loss function and followed by computing the correlation coefficient matrix [102]. We then apply a threshold on the correlation coefficient matrix to obtain a binary graph structure matrix. For testing, we multiply the features from the network trained on malignancy with the corresponding coefficient vector W to obtain the score.

For evaluation, we used metrics for both classification and regression. We calculated classification accuracy by considering classification to be successful if the predicted score lies in ± 1 of the true score. We also reported the average absolute score difference between the predicted score and the

true score.

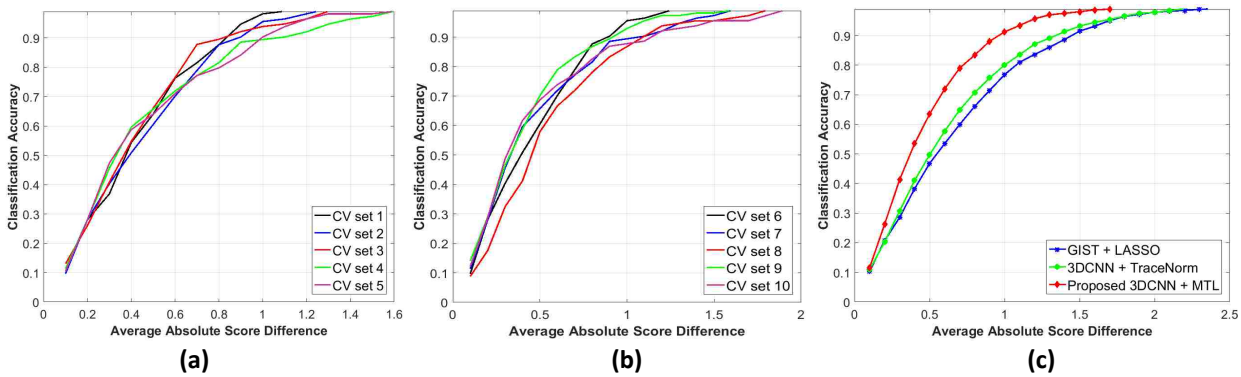


Figure 5.5: Plots to show classification accuracy against various threshold values for the average absolute score difference. The graphs in (a) and (b) represent results from 10 different cross-validation (CV) sets. It can be seen that the classification accuracy increases, as we increase the threshold value for the absolute score difference. The graph (c) shows the improved performance of the proposed method in comparison with GIST+LASSO and 3DCNN with Trace Norm.

Table 5.3 shows the comparison of our proposed Multi-task learning method with GIST features [105] +LASSO and 3D CNN Multi-task learning with trace norm. Our proposed graph regularized MTL outperforms the other methods with a significant margin. Our approach improves the classification accuracy over GIST features by about 15% and over trace norm regularization by 11%. Moreover, the average absolute score difference reduces by 32% and 27% when compared with GIST and trace norm respectively.

We also plotted classification accuracy against different thresholds for average absolute score difference. Figure 5.5 (a) and (b) show the plot on different cross-validation sets. It can be noticed that across different validation sets, the predicted malignancy scores of around 70% of the nodules lie within a margin of ± 0.6 which increases to around 90% when ± 1 margin is used. Figure 5.5 (c) shows the comparison with the other methods, where the proposed approach outperforms them over all values of average absolute score difference. Figure 5.6 shows the qualitative results from our proposed approach.

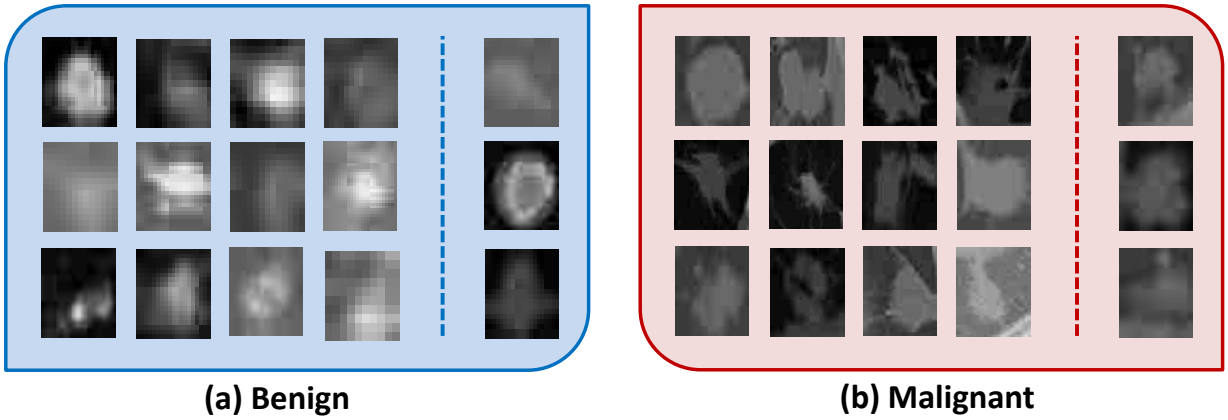


Figure 5.6: Qualitative results using our proposed approach. (a) and (b) show axial views of benign and malignant nodules respectively, where first four columns consist of successful cases (where prediction was within ± 1 of the expert score) and the last column (after dotted line) shows failure cases.

In order to evaluate the significance of transfer learning via fine-tuning, we project the features onto a low dimensional space. This is done by computing the proximity, between boundary points using t-distributed stochastic neighborhood embedding (t-SNE) [76]. As our feature space is high dimensional (4096-dimension), t-SNE is useful in revealing the structure of data at different scales. It can be seen in Figure 5.7 that fine-tuning the network on the lung nodule dataset distinctively improves the separation between benign and malignant classes.

5.2.3 Summary

In this work, we proposed a framework to stratify the malignancy of lung nodules using 3D CNN and graph regularized sparse multi-task learning. To the best of our knowledge, this is for the first time, transfer learning is employed over 3D CNN to improve lung nodule characterization. The task of data collection, especially in medical imaging fields, is highly regulated and the availability of experts for annotating these images is restricted.

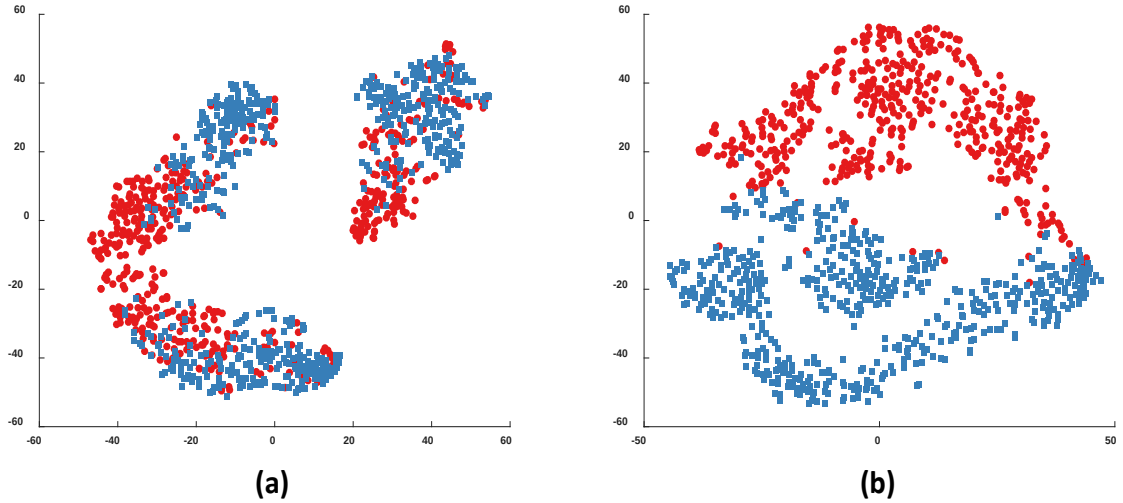


Figure 5.7: Effect of fine-tuning on 3D CNN features. *t*-SNE visualization for features obtained from (a) pre-trained network and (b) network after fine-tuning. The separation between features belonging to two classes, i.e. benign nodules (represented in blue) and malignant nodules (shown in red) can be readily observed in (b).

In this scenario, leveraging the availability of crowdsourced and annotated data such as user captured videos can be instrumental in training discriminative models. However, given the diversity in data from these two domains (i.e. medical and non-medical user collected videos), it is vital to perform transfer learning from source domain (user collected videos) to the target domain (medical imaging data). To establish this observation and to visualize features, we used t-SNE to project high dimensional features onto a low dimensional space (2D space), where the separation between classes was evident in the case of transfer learning.

Moreover, in this work, we also empirically explored the importance of high-level nodule attributes such as calcification, sphericity, lobulation and others to improve malignancy determination. Rather than manually determining these attributes we used 3D CNN to learn discriminative features corresponding to these attributes. The 3D CNN based features from these attributes are fused in a graph regularized sparse multi-task learning.

Another important imaging modality for lung nodule diagnosis is Positron Emission Tomography (PET). It has been found that the combination of PET and CT can improve the diagnostic accuracy of solitary lung nodules [106]. With the increase in the availability of PET/CT scanners, our future work will involve their utilization for simultaneous detection and characterization of pulmonary nodules.

5.3 Deep Multi-modal Diagnosis of IPMN

In this section, we present a computer-aided diagnosis system for the classification of IPMN in multi-modal MRI images. Here the goal is to exploit the complementary information from T1 and T2 weighted MRI scans and to utilize them for the improved diagnosis. We present a canonical correlation based framework to obtain a joint feature representation which is then used to obtain the diagnostic decision.

5.3.1 CNN for Multi-modal Feature Representation:

Problem Formulation:

Our proposed approach consists of inputs from two different MRI image modalities T1 and T2. Let $\mathcal{I}_{T1} = [I_1, I_2 \dots I_{N_1}] \in \mathbb{R}^{X_1 \times Y_1 \times N_1}$ be the T1 scan whereas the corresponding T2 scan is represented as $\mathcal{J}_{T2} = [J_1, J_2 \dots J_{N_2}] \in \mathbb{R}^{X_2 \times Y_2 \times N_2}$, with N_1 and N_2 number of slices.

Consider I_u be the slice with pancreas from T1 scan and J_v be the slice from T2. Predicting the label from a single slice, however, may yield to hypersensitivity in annotation labels as well as miss important contextual information from the other slices. In order to address these issues, we sample $k = 2$ consecutive slices before and after I_u and J_v. Since the input to the deep network is a 2D

image, we use Maximum and Minimum Intensity Projections to combine information across various slices into a single slice. We employ minimum intensity projection for T1 scans since IPMN and pancreatic cysts are *hypo-intensity* regions in T1 scans. In contrast, we use maximum intensity projection for T2 scans, because IPMN and pancreatic cysts correspond to *hyper-intensity* regions in these scans. The intensity projections corresponding to T1 and T2 scans can be represented as:

$$\begin{aligned} \mathcal{I}' &= \min_{Z_1} [\mathcal{I}(X_1, Y_1, Z_1)], \\ \mathcal{J}' &= \max_{Z_2} [\mathcal{J}(X_2, Y_2, Z_2)], \end{aligned} \tag{5.12}$$

where Z_1 and Z_2 consists of k slices around I_u and J_v respectively. Moreover, $\mathcal{I}'$ and $\mathcal{J}'$ represent the intensity projections from T1 and T2 scans, respectively. The overview of the proposed approach is shown in Figure 5.8

Network Architecture:

In order to obtain deep feature representation for our proposed IPMN classification approach, we use (fast) CNN-F architecture trained on ImageNet [71]. The architecture consists of 5 convolutional and 3 fully-connected layers. The input 2D image is resized to 224 $\times$ 224. The first convolutional layer contains 64 filters with stride 4 and there are 256 filters with stride 1 in the other 4 convolutional layers. Our input to the network is the 2D intensity projections $\mathcal{I}'$ and $\mathcal{J}'$, whereas the features are extracted from the second fully connected layer without applying non-linearities such as ReLU (Rectified Linear Units). The features are ℓ_2 normalized to obtain the final representation.

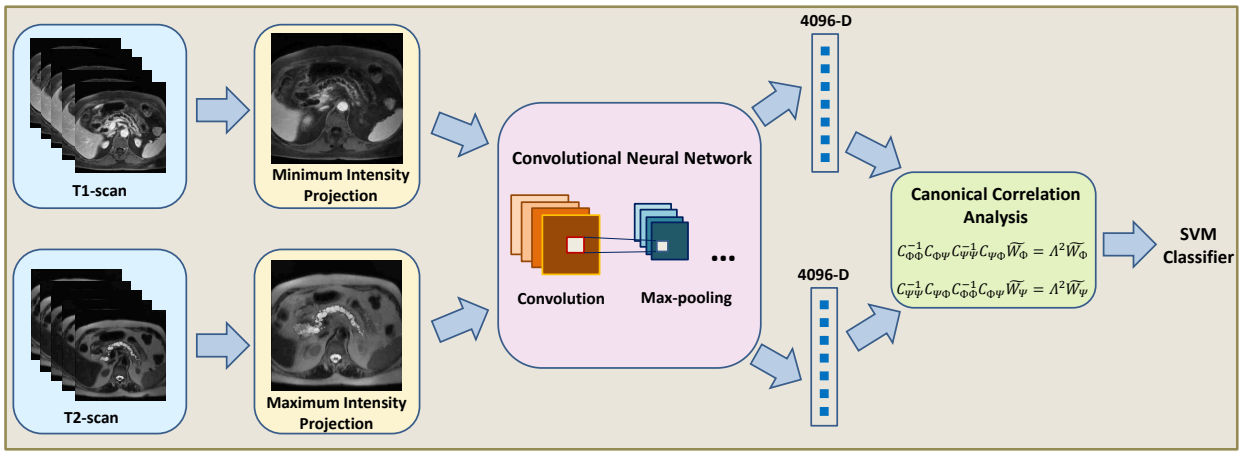


Figure 5.8: An overview of the proposed method. First, the minimum and maximum intensity projections are computed corresponding to T1 and T2 scans respectively. The intensity projections are then fed into a pre-trained Convolutional Neural Network (CNN) to obtain feature representation. Canonical Correlation Analysis (CCA) based feature fusion is performed in order to obtain discriminative and transformed feature representation. Finally, an SVM based classifier is employed to obtain the final label (normal or IPMN).

5.3.2 Feature Fusion with Canonical Correlation Analysis:

The next step is to combine information from the two imaging modalities so as to improve the classification performance. As these two imaging modalities (T1 and T2) have complementary information, the fusion of features from these modalities can help improve IPMN diagnosis. Assume that $\Phi \in \mathbb{R}^{n\times p}$ and $\Psi \in \mathbb{R}^{n\times q}$ comprise the deep features from the intensity projections of n training images from T1 and T2 scans respectively. Each sample has a corresponding binary label given by $\mathcal{Y} = [y_1, y_2 \ldots y_n]$, where $\mathcal{Y} \in \{0,1\}^{n\times 1}$. Consider $C_{\Phi\Phi} \in \mathbb{R}^{p\times p}$ and $C_{\Psi\Psi} \in \mathbb{R}^{q\times q}$ represent the within sets covariance matrices of Φ and Ψ respectively. Additionally, the between set covariance matrix is referred as $C_{\Phi\Psi} \in \mathbb{R}^{p\times q}$ such that $C_{\Psi\Phi} = C_{\Phi\Psi}^T$. The covariance matrix $\mathcal{C}$

can therefore be written as:

$$\mathcal{C} = \begin{pmatrix} cov(\Phi) & cov(\Phi, \Psi) \\ cov(\Psi, \Phi) & cov(\Psi) \end{pmatrix} = \begin{pmatrix} C_{\Phi\Phi} & C_{\Phi\Psi} \\ C_{\Psi\Phi} & C_{\Psi\Psi} \end{pmatrix} \tag{5.13}$$

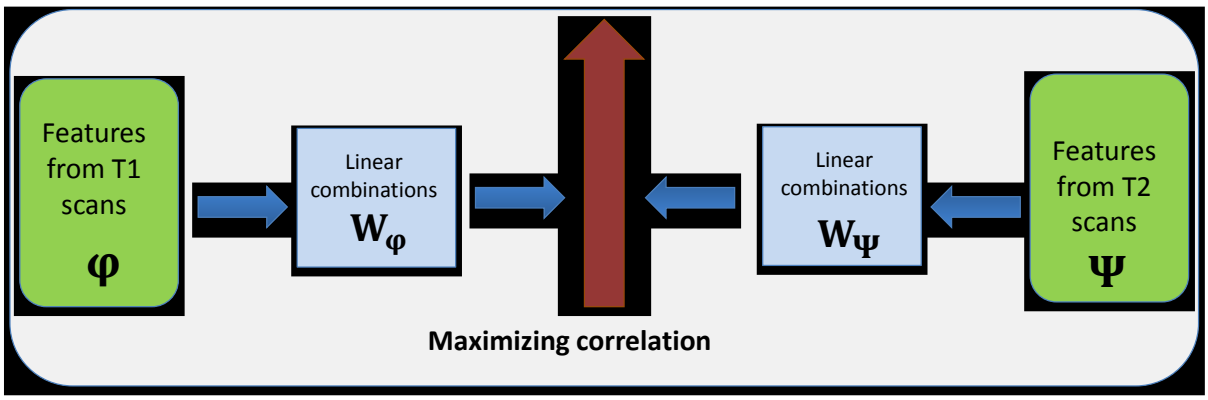


Figure 5.9: An illustration to show CCA computed between features from two different imaging modalities i.e. T1 and T2 MRI scans

In this regard, CCA is employed to find the linear combinations, $\Phi^* = W_\Phi^T \Phi$ and $\Psi^* = W_\Psi^T \Psi$ such that the pair-wise correlation between the two sets is maximized [107]. CCA is a method that can help explore the relationship between the two multi-variate variables. The pairwise correlation between the two sets can be modeled as:

$$corr(\Phi^*, \Psi^*) = \frac{cov(\Phi^*, \Psi^*)}{var(\Phi^*).var(\Psi^*)}, \tag{5.14}$$

where $cov(\Phi^*, \Psi^*) = W_\Phi^T C_{\Phi\Psi} W_\Psi$, $var(\Phi^*) = W_\Phi^T C_{\Phi\Phi} W_\Phi$ and $var(\Psi^*) = W_\Psi^T C_{\Psi\Psi} W_\Psi$. The covariances are then used to find the transformation matrices W_Φ and W_Ψ using the following equations:

$$
\begin{aligned}
C_{\Phi\Phi}^{-1} C_{\Phi\Psi} C_{\Psi\Psi}^{-1} C_{\Psi\Phi} \widetilde{W}_{\Phi} &= \Lambda^2 \widetilde{W}_{\Phi}, \\
C_{\Psi\Psi}^{-1} C_{\Psi\Phi} C_{\Phi\Phi}^{-1} C_{\Phi\Psi} \widetilde{W}_{\Psi} &= \Lambda^2 \widetilde{W}_{\Psi}.
\end{aligned} \tag{5.15}
$$

In the above equation, $\widetilde{W}_{\Phi}$ and $\widetilde{W}_{\Psi}$ are the eigenvectors and Λ^2 is the eigenvalue diagonal matrix.

Lastly, the final feature matrix can be represented as the sum of the transformed feature matrices from the two modalities:

$$
F = W_{\Phi}^T \Phi + W_{\Psi}^T \Psi = \begin{pmatrix} W_{\Phi} \\ W_{\Psi} \end{pmatrix}^T \begin{pmatrix} \Phi \\ \Psi \end{pmatrix}. \tag{5.16}
$$

The learned transformation is also applied to the features from test images in order to obtain the final transformed testing features. An illustration to show CCA computed from T1 and T2 scans is depicted in Figure 5.9.

5.3.3 Materials

We evaluated our proposed approach for the classification of IPMN on a dataset comprising post-contrast volumetric T1 and T2 MRI scans from 139 subjects. The scans were labeled as normal or IPMN using pathology report obtained after surgery. Out of 139 scans, 108 were from subjects diagnosed with IPMN, whereas the rest of the 31 subjects were normal. The in-plane spacing (xy-plane) of T1-weighted scans was ranging from 0.66 mm to 1.48 mm and that of T2-weighted scans from 0.47 mm to 1.41 mm. For pre-processing, we first apply N4 bias field correction [108] to each scan in order to minimize intensity inhomogeneity. Next, we use a curvature anisotropic

image filter to smooth images while preserving edge information. For each image, a single slice which has a significant portion of the pancreas is annotated to be normal or IPMN.

5.3.4 Experiments and Results

In order to account for the misalignment between T1-weighted and T2-weighted scans, we performed Multi-resolution image registration using image pyramids [109]. The registration results were examined and images with mis-registration were removed from the final evaluation set. Our final evaluation set comprised 139 scans from each modality and we performed 10 fold cross validation over the dataset. The minimum (maximum) intensity projection images from T1 (T2) scans were fed into the deep CNN-F network and feature representation from each of these images was used to obtain the final CCA based discriminative representation (Eq. 5.16). We then employed Support Vector Machine (SVM) classifier to obtain the final classification labels i.e. normal vs IPMN.

Table 5.4: Results for accuracy, sensitivity and specificity of the proposed multi-modal fusion approach along with standard error of the mean (SEM) in comparison with single modality and feature concatenation based approaches.

Methods	**Accuracy** (SEM %)	**Sensitivity** (SEM %)	**Specificity** (SEM %)
T1-weighted	84.23 (1.10)	89.16 (0.88)	55.00 (4.16)
T2-weighted	61.04 (1.35)	59.59 (2.25)	57.67 (2.95)
Concat. of T1 & T2	82.09 (1.01)	88.49 (0.90)	49.33 (3.48)
Feature Fusion (Proposed)	**82.80 (1.17)**	**83.55 (1.13)**	**81.67 (2.53)**

We compared our proposed multi-modal feature fusion based approach with single modality and feature concatenation based approaches. Since there exists an imbalance between the number of positive and negative examples, we performed Adaptive Synthetic Sampling (ADASYN) [110] to

generate synthetic samples. This sampling enabled to generate synthetic feature examples from the minority class (normal).

Table 5.4 shows the results of our proposed approach in varying conditions. It can be observed that the performance of our proposed approach significantly outperforms the single modality and feature concatenation based approaches. The T1 based classification yielded the highest sensitivity, but with very low *specificity*. For IPMN classification, low specificity can be a serious problem as that can lead to unwarranted surgery and resection. In this regard, our proposed approach reports more than **30% improvement** in specificity in comparison with the feature concatenation based approach.

Table 5.5: Classification accuracy along with standard error of the mean (SEM) for three class classification (normal, low grade IPMN and high grade IPMN) of proposed approach in comparison with other approaches.

Methods	**Accuracy % (SEM %)**
T1-weighted	58.30 (1.52)
T2-weighted	45.93 (1.60)
Concat. of T1 and T2	56.81 (1.51)
Feature Fusion (Proposed)	**64.67 (0.83)**

It is important to note that since our proposed approach is based on the correlation and covariance in the data, it doesn't require explicit sample balancing using ADASYN. Moreover, for experiments, we also tried features from various layers of CNN-F as well as features from deeper residual networks such as ResNet-50, ResNet-101, and ResNet-152 [111]. The best classification results, however, were obtained using the second fully-connected layer of CNN-F. Figure 5.10 shows the qualitative results of our proposed approach with intensity projections from T1 and T2 scans. The cases shown in green are the correctly classified as IPMN whereas those shown in red are incorrectly classified as normal.

Low and High grade IPMN classification:

We also performed a 3-class classification using our proposed approach. Out of 108 IPMN subjects, 48 had low-grade IPMN whereas the remaining 60 had high-grade IPMN or invasive carcinoma. Using the features obtained from the CCA based fusion, we train a 3-class SVM classifier with classes normal, low-grade IPMN and high-grade IPMN. These diagnostic labels were obtained using the pathology report after surgery. Table 5.5 shows the performance of our proposed approach for normal, low-grade and high-grade IPMN classification.

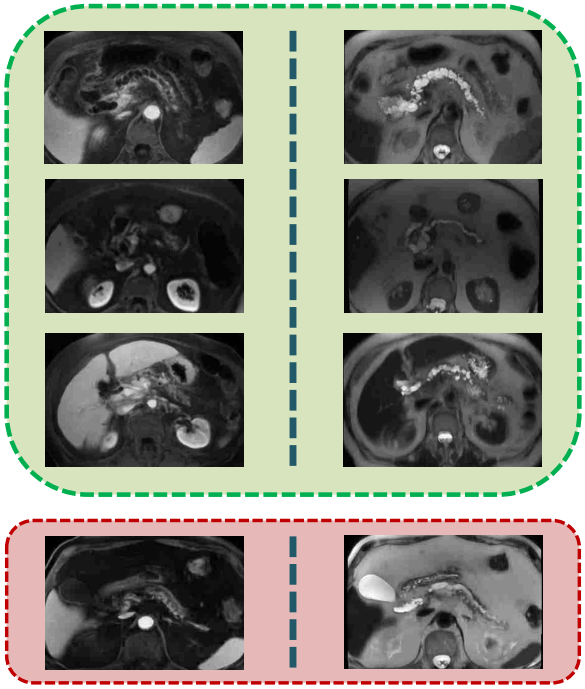


Figure 5.10: Qualitative results of our proposed approach, showing minimum and maximum intensity projected images for T1 and T2 scans on left and right respectively. Each row represents a different case where the images correctly classified as IPMN are shown in green, whereas the misclassification of IPMN as normal are shown in red.

The proposed CCA based classification approach outperforms single modality and feature concatenation based approaches. The CCA based approach reports around **8% improvement** in com-

parison to the feature concatenation based approach.

5.3.5 Summary

Pancreatic cancer is projected to become the second leading cause of cancer-related deaths before 2030 [112]. IPMNs are the radiographically identifiable precursor to pancreatic cancer. In this chapter, we proposed a multi-modal feature fusion framework to perform the classification of IPMN. Inspired by the clinical need to identify subjects with IPMN, our proposed approach can help radiologists in diagnosing invasive pancreatic carcinoma. In contrast to previous studies, this is the first approach to use deep CNN feature representation for IPMN diagnosis. Moreover, we empirically show the importance of feature level fusion of two different MRI imaging modalities i.e. T1 and T2 scans.

Another advantage of our proposed approach is that it doesn't require manual segmentation of pancreas or cysts as in other approaches. We only need to identify a single slice where pancreatic tissues can be prominently observed. Additionally, by using the intensity projections across a consecutive set of slices, we can develop robustness to the manual selection of a single slice. As the CCA is used to learn the transformation, its use also circumvents the need to have explicit data balancing in the cases of imbalance between positive and negative examples.

As an extension to this study, our future work will involve joint detection and diagnosis of IPMN in MRI scans. As the number of subjects undergoing screening for IPMN increases, we can get sufficient data to perform an end-to-end training or fine-tuning of a 3D convolutional neural network. The use of Generative Adversarial Networks (GANs) [113] can assist in data augmentation by generating realistic examples to further improve the training of the network.

Furthermore, the segmentation of the pancreas and IPMN cysts can help in localizing the regions of

interest. These regions can be used not only to extract discriminative imaging features but also to extract important measurements such as the diameters of the main pancreatic duct and cysts [28]. The inclusion of additional imaging modalities such as CT scans along with demographic and clinical characteristics, including age, gender, family history, symptoms and body fat can help further improve diagnostic decision making in the future.

CHAPTER 6: UNSUPERVISED TUMOR CHARACTERIZATION

Related Publication:

Supervised and Unsupervised Tumor Characterization in the Deep Learning Era. Sarfaraz Hussein, M.M. Chuquicusma, P. Kandel, C.W. Bolan, M.B. Wallace, and U. Bagci. IEEE Transactions on Medical Imaging (IEEE-TMI; Under Revision)

In this chapter, we focus on the characterization of lung and pancreatic tumors with unsupervised learning approaches. In contrast to the work in previous chapters which addressed the classification problem with supervised approaches, this work discusses tumor classification with unsupervised learning. Generally, the classification tasks in the visual domain are dealt with supervised approaches where a large number of labeled examples are available. However, in cases where obtaining labels for training samples is not feasible or practical, unsupervised learning strategies can be applied. Conventionally, unsupervised approaches perform inferior to supervised approaches, which is intuitive. In this case, the presence of any complementary information can be vital which may help in improving the performance.

Since the task of annotating medical images is laborious, expensive and time-consuming, we explore the potential of unsupervised learning approaches for developing CAD systems. Our proposed unsupervised framework includes three steps (Figure 6.1). Given the unlabeled set of images, we first perform clustering on the appearance features so as to obtain an initial set of class labels. These initial labels are progressively refined in the next steps. In the second step, we compute label proportions corresponding to each cluster using the obtained initial labels. In the absence of instance level labels to train a classifier, we utilize bag level labels and label proportions. In this

regard, we finally train a classifier using the label proportions and clusters to obtain the final class labels.

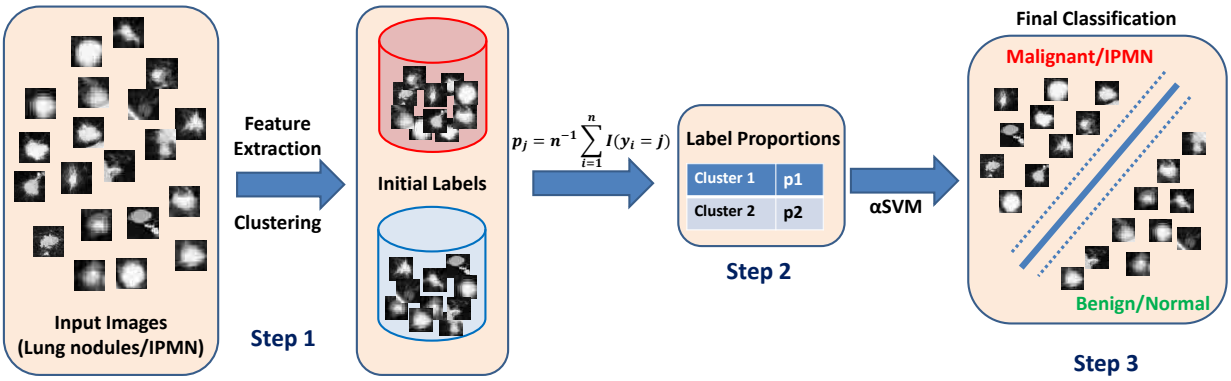


Figure 6.1: An outline of the proposed unsupervised approach. Given the input images, we compute GIST features and perform k-means clustering to get the initial set of labels which can be noisy. Using the set of labels, we compute label proportions corresponding to each cluster/group (Eq. 6.4). We finally employ $\propto$SVM to learn a discriminative model using the features and label proportions.

6.1 Initial Label Estimation

Let $X = [x_1, x_2 \ldots x_n] \in \mathbb{R}^{n \times d}$ represents the input matrix which contains features from n images such that $x \in \mathbb{R}^d$. In the absence of labels corresponding to examples, we need to obtain an initial set of labels with supervision. In order to obtain an initial set of labels corresponding to each sample, we cluster the data into $2 \leq k < n$ clusters using k-means clustering. Let M represents $|X| \times k$ membership matrix which denotes the membership assignment of each sample to a cluster. The optimal clustering would minimize the following objective function:

$$\underset{M}{\operatorname{argmin}} \sum_{j=1}^{k} M(i,j) \left\| x_i - \mu_j \right\|^2 , \tag{6.1}$$

where μ_j is the mean of the samples in cluster j. The label y_i for the ith sample can then be

estimated as:

$$y_i = \underset{j}{\operatorname{argmax}}\, M(i,j) \tag{6.2}$$

These labels serve as an initial set so it is, therefore, reasonable to assume that the output labels from clustering are noisy. These initial labels are used to estimate label proportions which are then used to train proportion-SVM ($\propto$SVM) for further improved classification.

6.2 Learning with the Estimated Labels

Since our initial label estimation approach is unsupervised, there are uncertainties associated with them. It is, therefore, reasonable to assume that learning a discriminative model based on these noisy instance level labels can deteriorate classification performance. In order to address this issue, we model the instance level labels as *latent* unknown variables and thereby consider group/bag level labels.

Inspired by $\propto$SVM approach [114], which models the latent instance level variables using the known group level label proportions, we formulate our learning problem such that clusters are analogous to the groups. In our formulation, each cluster j can be represented as a group such that the majority of samples belong to the class j. Considering the groups to be disjoint such that $\bigcup_{j=1}^{k}\Omega_j = 1, 2, \ldots n$, and Ω represents groups; the objective function of the large-margin $\propto$SVM after convex relaxation can be formulated as:

$$\begin{aligned} &\min_{\mathbf{y}\in\mathcal{Y}} \min_{w} \left(\frac{1}{2} w^T w + C \sum_{i=1}^{n} L(y_i, w^T \phi(x)) \right) \\ &\mathcal{Y} = \left\{ \mathbf{y} \,\middle|\, |\widetilde{p_j}(\mathbf{y}) - p_j| \leq \epsilon, y_i \in \{-1, 1\}\, \forall_{j=1}^{k} \right\}, \end{aligned} \tag{6.3}$$

where $\widetilde{p}$ and p represent the estimated and true label proportions, respectively. In Eq. 6.3, $\mathbf{y}$ is the set of instance level labels, $\phi(.)$ is the input feature, C denotes cost parameter and $L(.)$ represents the hinge-loss function for maximum-margin classifiers such as SVM.

The optimization in Eq.6.3 is, in fact, an instance of Multiple Kernel Learning, which can be solved using the cutting plane method where the set of active constraints is incrementally computed. The goal is to find the most violated constraint, which is further relaxed by aiming for any violated constraint.

6.3 Calculating Label Proportions

In the conventional $\propto$SVM approach, the label proportions are known a priori. Since our approach is unsupervised, both instance level labels and group label proportions are unknown. Moreover, establishing strong assumptions about the label proportions may affect learning. It is, however, reasonable to assume that a large number of instances in any group carry the same label and there may be a small number of instances which are outliers. We, therefore, assign the group level label based on the majority of the examples present in that group. In order to determine the label proportions in a data-driven manner, we use the estimated labels obtained from clustering. The label proportion p_j corresponding to the group j can be represented as:

$$p_j = n^{-1} \sum_{i=1}^{n} I(y_i = j), \tag{6.4}$$

where I(.) is the indicator function which yields 1 when $y_i = j$. The $\propto$SVM is trained using the image features and label proportions to classify the testing data. It is important to mention that since our approach is unsupervised we don't use the ground truth labels for training rather they are only used for evaluations. In addition, clustering and label proportion calculation are only

performed on the training data and the testing data remains completely unseen for $\propto$SVM.

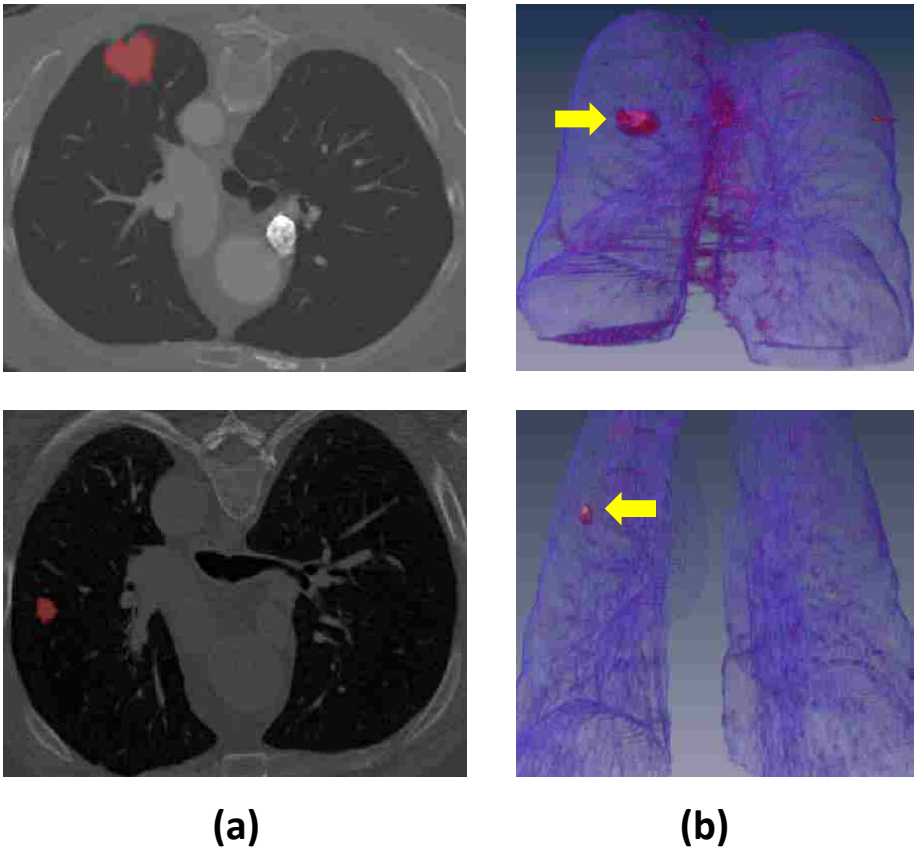


Figure 6.2: Representation of lung nodules in (a) axial CT slices and in (b) volume renderings. The top row shows malignant nodule where the second row depicts an instance of benign nodule. The nodules are highlighted in red in (a) and are pointed by arrows in (b).

6.4 Experiments and Results:

For unsupervised learning, evaluations were performed on both lung nodules and IPMN datasets. Illustrations representing lung nodule and IPMN are shown in Figure 6.2 and Figure 6.3 respectively. For lung nodules, we used LIDC-IDRI dataset with the same specifications as used in Chapter 5. The data for IPMN contains T2 MRI axial scans from 171 subjects. The scans were

labeled by a radiologist as normal or IPMN. Out of 171 scans, 38 subjects were normal, whereas the rest of 133 were from subjects diagnosed with IPMN. For experiments related to IPMN, 2D axial slices with pancreas (and IPMN) were cropped to generate Region of Interest (ROI) as shown in Figure 6.3.

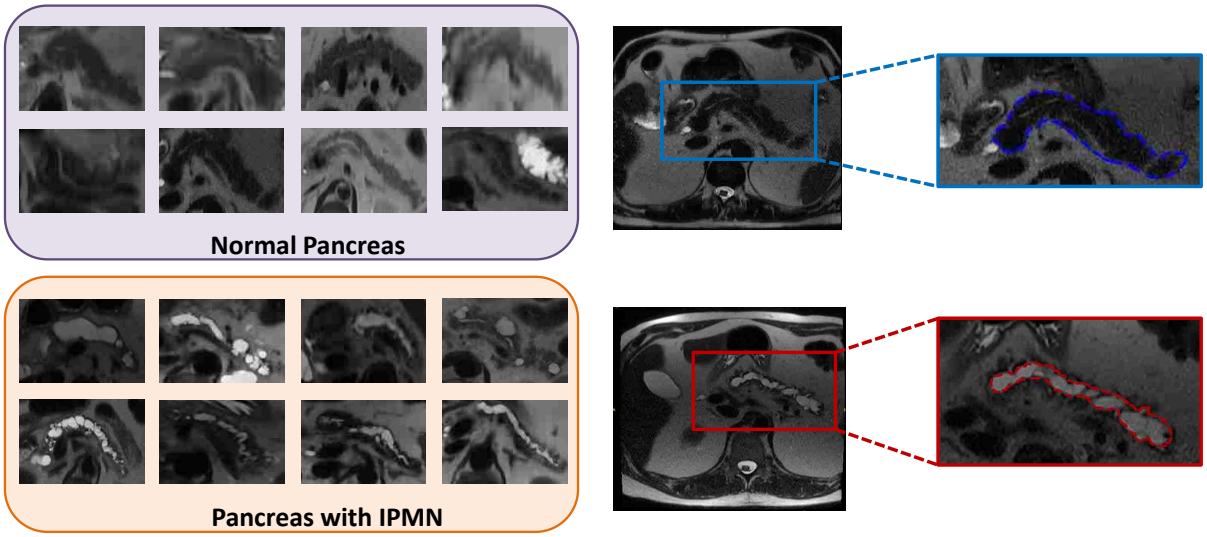


Figure 6.3: Axial T2 MRI scans illustrating pancreas. The top row shows different ROIs of pancreas, along with a magnified view of a normal pancreas (outlined in blue). The bottom row shows ROIs from subjects with IPMN in the pancreas, which is outlined in red.

The challenge of diagnosing IPMN and large intra-class variation, especially due to varying shapes of the pancreas can also be observed in Figure 6.3.

In order to compute image level features, we used GIST descriptors [105]. The number of clusters is fixed as 2, which accounts for benign and malignant classes. The clustering result was checked to assign benign and malignant labels to the clusters. We used 10 fold cross-validation to evaluate our proposed approach. The training samples along with the label proportions generated using clustering served as the input to $\propto$SVM with a linear kernel.

To evaluate our unsupervised approach we used accuracy, sensitivity and specificity as metrics. It

can be observed in Table 6.1 that the proposed combination of clustering and $\propto$SVM significantly outperforms other approaches in accuracy and sensitivity. In comparison with clustering+SVM, the proposed framework yields almost 21% improvement in sensitivity for lung nodules and around 7% improvement for IPMN classification. The low sensitivity and high specificity of clustering and clustering+SVM approaches can be explained by disproportionate assignment of instances as benign (normal) by these two approaches, which is not found in the proposed approach. At the same time, the proposed approach records around 24% and 9% improvement in accuracy as compared to clustering for lung nodules and IPMN, respectively.

Table 6.1: Average classification accuracy, sensitivity and specificity of the proposed *unsupervised* approach for IPMN and lung nodule classification with other methods

Evaluation Set	Methods	Accuracy	Sensitivity	Specificity
IPMN Classification	Clustering	49.18%	45.34%	**62.83%**
	Clustering + SVM	52.03%	51.96%	50.5%
	Proposed approach	**58.04%**	**58.61%**	41.67%
Lung Nodule Classification	Clustering	54.83%	48.69%	60.04%
	Clustering + SVM	76.04%	57.08%	**91.28%**
	Proposed approach	**78.06%**	**77.85%**	78.28%

6.4.1 Are Deep Features good for Unsupervised Classification?

Given the success of deep learning features for image classification tasks and their popularity with the medical imaging community, we explored their performance to classify lung nodules and IPMN in an unsupervised manner. For this purpose, we used a pre-trained deep CNN architecture to extract features and then perform clustering to obtain baseline classification performance. We extracted features from fully-connected layers 7 and 8 of Fast-VGG [71] with and without apply-

ing ReLU non-linearity. Classification accuracy, using clustering over these features is shown in Figure 6.4.

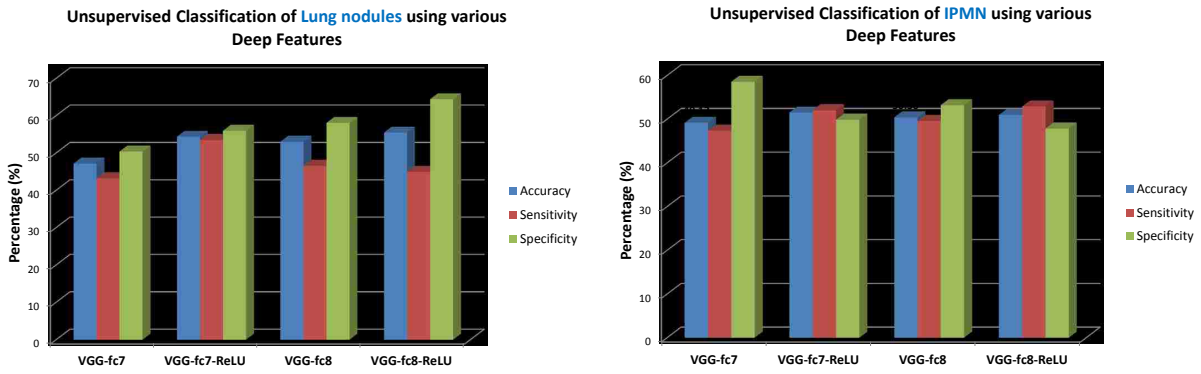


Figure 6.4: Influence of deep learning features obtained from different layers of a VGG network with and without ReLU non-linearities. The graph on the left shows accuracy, sensitivity and specificity for unsupervised lung nodule classification (clustering), whereas the right one shows the corresponding results for IPMN.

It can be seen in Figure 6.4 that the features with non-linearity (ReLU) are more discriminative for classification using clustering as compared to without ReLU. The same trend can be observed for both lung nodules and IPMN classification using VGG-fc7 and VGG-fc8 layers. Owing to the larger evaluation set, the influence of ReLU is more prominent for lung nodules as compared to IPMN. Although the results between VGG-fc7 and VGG-fc8 are not substantially different, highest accuracy for IPMN can be obtained by using VGG-fc7-ReLU features and for lung nodules by using VGG-fc8-ReLU features. The non-linearity induced by ReLU clips the negative values to zero, which can sparsify the feature vector and can reduce overfitting.

Table 6.2: Classification of IPMN and Lung Nodules using different features and supervised learning classifiers.

Evaluation Set	Features	Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
IPMN Classification	GIST	SVM	76.05	83.65	**52.67**
		RF	81.9	93.69	43.0
	VGG-fc7	SVM	84.18	96.91	44.83
		RF	81.96	94.61	42.83
	VGG-fc8	SVM	**84.22**	**97.2**	46.5
		RF	80.82	93.4	45.67
Lung Nodule Classification	GIST	SVM	81.56	71.31	90.02
		RF	81.64	76.47	**85.97**
	VGG-fc7	SVM	77.97	75.2	80.6
		RF	**81.73**	**78.24**	84.59
	VGG-fc8	SVM	78.76	74.67	82.29
		RF	80.51	76.03	84.24

6.4.2 Classification using Supervised Learning

In order to establish the upper-bound on the classification performance, we trained linear SVM and Random Forest using GIST and different deep learning features with ground truth labels on the same 10 fold cross-validations sets.

Table 6.2 lists the classification accuracy, sensitivity, and specificity using GIST, VGG-fc7 and VGG-fc8 features for both IPMN and lung nodules. For both VGG-fc7 and VGG-fc8, we used features after ReLU since they are found to be more discriminative (Figure 6.4). Interestingly, for lung nodules, VGG-fc7 features along with RF classifier are reported to have comparable results to the combination of GIST and RF classifier. This can be explained by the fact that deep networks are pre-trained on ImageNet dataset as compared to handcrafted features such as GIST, which don't require any training. On the other hand, for smaller datasets such as IPMN, deep features are found to perform better as compared to GIST. In order to balance the number of positive (IPMN) and negative (normal) examples, which can be a critical drawback otherwise, we performed Adap-

tive Synthetic Sampling [110]. This was done to generate synthetic examples in terms of features from the minority class (normal).

6.4.3 Generative Adversarial Networks (GAN)

Recently, generative adversarial networks (GANs) [113] have gained considerable attention in deep learning and have been applied for different applications in visual domain [115, 116] including but not limited to image super-resolution [117], inpainting [118], video generation [119] as well as non-imaging clinical tasks such as molecule development in oncology [120]. GANs are utilized to generate realistic images and videos which can be both class dependent or class agnostic. GANs can capture meaningful variation in the data which can be important for data augmentation and data compression. GAN consists of a generator and discriminator, where the goal of the generator is to generate as realistic samples as possible whereas the discriminator aims to successfully differentiate between the original samples and those produced by the generator. Let **G** be the generator, which aims to learn a distribution, p_z from noise, and **D** be a discriminator, which sifts real data (sampled from true distribution $p_{data}(x)$) from the artificial data (output of the generator). The two-player minmax game between **G** and **D** is modeled with the following value function:

$$\begin{aligned} \min_{\mathbf{G}} \max_{\mathbf{D}} V(\mathbf{D}, \mathbf{G}) = \mathbb{E}_{x \sim p_{data}(x)}[log(D(x))] \\ + \mathbb{E}_{z \sim p_z(z)}[log(1 - D(G(z)))], \end{aligned} \tag{6.5}$$

where $\mathbb{E}$ represents expectation.

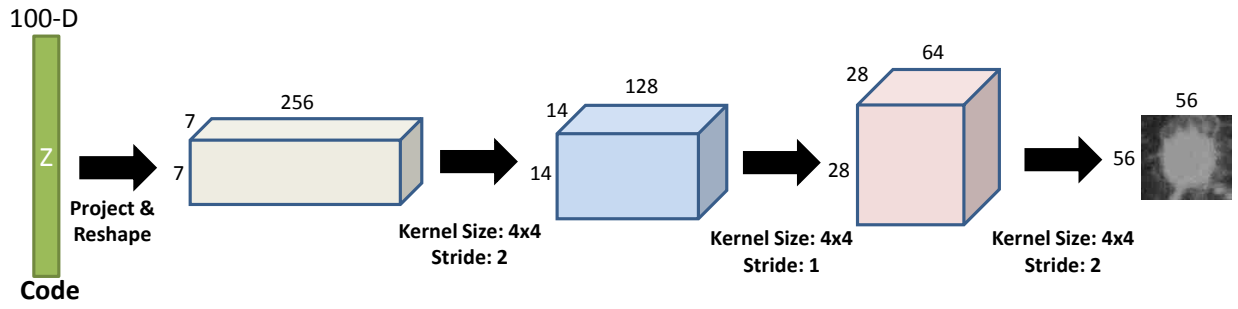


Figure 6.5: The generator architecture of DCGAN used to perform unsupervised representation learning.

In our work, we use Deep Convolutional GAN (DC-GAN) [115] to perform unsupervised representation learning and then use those representations to perform supervised classification. DC-GANs are an extension of standard GAN by having convolutional layers along with batch normalization, removing fully connected layers, and utilizing ReLU, leaky ReLU, and tanh activation functions. Our DC-GAN architecture is shown in Figure 6.5. The generator (**G**) comprises 3 deconvolution filters of sizes 128, 64 and 1 respectively. The input to G is a 100-dimensional vector which initially consists of noise. Consequently, the initial output of the generator is noise. However, as the training progresses and the discriminator correctly classifies the generated samples as fake, the generator starts improving and producing realistic images. The discriminator consists of 2 convolutional layers of sizes 32 and 64 followed by a fully-connected layer of dimension 3136. The features from there FC layer of D are used to train off-the-shelf classifiers such as SVM and RF. The classification results with features from DC-GAN are shown in Table 6.3. It can be seen that the unsupervised feature representation with classifiers reports comparable results to those obtained from the pre-trained network (Table 6.2) for IPMN classification. For lung nodules, classification using pre-trained networks such as VGG outperforms classification over DC-GAN features. One potential reason for this observation can be instability in training of GAN and mode-collapse. In a semi-supervisory scheme, the discriminator of the GANs can be made to output class labels [121].

Table 6.3: Supervised classification for IPMN and Lung Nodules using unsupervised feature learning via Deep Convolutional Generative Adversarial Networks (DC-GAN)

Evaluation Set	Methods	Accuracy	Sensitivity	Specificity
IPMN Classification	SVM	**83.66**	**96.36**	42.83
	RF	81.37	93.92	**44.50**
Lung Nodule Classification	SVM	75.09	68.31	80.41
	RF	**77.96**	**72.53**	**82.45**

6.5 Summary

In this chapter, we explored the potential of unsupervised learning for lung nodule and IPMN classification. Since in most medical imaging tasks radiologists are required to get annotations, acquiring labels to learn machine learning models is more cumbersome and expensive as compared to other computer vision tasks. In order to address this challenge, we employed clustering to obtain an initial set of labels. Since the instance level labels obtained from clustering may not be reliable enough to train a discriminative classifier, we used label proportions obtained from clustering to train a $\propto$SVM. Our proposed combination of clustering and $\propto$SVM outperformed the other methods in evaluation metrics.

Following up on the application of deep learning in almost all tasks in the visual domain, we studied the influence of different pre-trained deep networks for lung nodule classification. We found that the GIST features were superior in performance as compared to the features from a pre-trained network. This observation can be explained by the fact that the deep networks were trained on ImageNet classification tasks so the filters in CNN were more tuned toward the nuances in natural images as compared to medical images. Moreover, in order to set a performance upper bound, we also evaluated supervised learning using GIST features and SVM. Finally, we employed Generative Adversarial Networks for unsupervised feature learning followed by the classification

using off-the-shelf classifiers. This work can be extended to incorporate unsupervised learning within the deep learning framework for tumor classification.

The future prospects of using different architectures to improve GAN training are promising. Instead of using hand-engineered priors of sampling in the generator, the work in [122] learned priors using denoising auto-encoders. For measuring the sample similarity for complex distributions such as those in the images, [123] jointly trained variational autoencoders and GANs. Moreover, the applications of CatGAN [124] and InfoGAN [125] for semi-supervised and unsupervised classification tasks in medical imaging are worth exploring as well.

CHAPTER 7: CONCLUSION AND FUTURE DIRECTIONS

This dissertation focuses on facilitating the exploration of a quantitative relationship between obesity and cancers. It proposes approaches for the segmentation and quantification of different types of adipose tissues (fat) along with the risk characterization of lung and pancreatic cancers.

7.1 Obesity

With regards to obesity, we proposed an automatic approach for the segmentation and quantification of white adipose tissue in CT scans. We first automatically identified abdominal region from whole body scan with deep learning features and one-shot learning. Given the abdominal region (region of interest) we first found the total adipose tissue using the intensity characteristics of the CT image. We then roughly marked the boundary separating subcutaneous and visceral fat using gradient cues. The initial boundary was progressively refined with geometric and appearance cues. We proposed to use median absolute deviation on the distances between the centroid of the abdominal cavity the SAT-VAT boundary. In the subsequent step, we proposed to compute local outlier score using the appearance features corresponding to the boundary points. As the last step, 3D sparse Conditional Random Fields were employed in order to fuse the appearance and geometric cues so as to yield the final SAT-VAT boundary. We evaluated our proposed framework on a dataset comprising 151 CT scans which can be considered as the largest central obesity quantification study to date.

As the second part of our obesity analysis, we detected and segmented brown adipose tissue in PET/CT scans. After obtaining the total white adipose tissue from a CT image, we sampled foreground and background locations from the corresponding PET image that serve as seeds to perform

segmentation. To incorporate varying contributions from PET and CT images, we used PET-guided CT co-segmentation with random walk. Finally, we proposed a new probabilistic distance metric that combined total variation and cramér-von Mises to perform false positive removal. Our work can be regarded as the first computer-aided detection and segmentation approach for brown fat analysis. As brown fat can serve as a natural defense against obesity, automatic detection of brown fat can be quite important to study its role with regards to various cancers.

7.2 Tumor Characterization

With regards to tumor characterization, we proposed different supervised and unsupervised strategies for the diagnosis of lung nodules and IPMN. For the *supervised* characterization of lung nodules, we first proposed a 2D CNN based approach where we trained a network end-to-end so as to realize the full potential of a deep network. In order to generate sufficient samples to train the network, we performed data augmentation with rotation, scale and different variants of noise. Gaussian Process Regression (GPR) was applied on the features obtained from the trained network yielding final malignancy scores. We also analyzed the influence of high-level attributes in confluence with deep learning features for malignancy prediction.

As an extension to the 2D network, we proposed a 3D CNN based approach for malignancy determination of lung nodules. Instead of hard assigning the nodule attributes scores, our proposed method learned them in a multi-task learning framework. In order to evade the need to have a large number of labeled training examples to train a 3D network, we fine-tuned a pre-trained network on the lung nodules dataset. We also incorporated the inconsistency in scoring from different radiologists within the framework. With multi-task learning, we aimed to find a joint feature representation from high-level nodule attributes so as to improve malignancy prediction.

We proposed a deep multi-modal approach for the automatic diagnosis of IPMN. Using the intensity projections, we obtained a 2D representation corresponding to 3D scans. Features using deep networks were extracted for intensity projections which were then fused with canonical correlation analysis (CCA), generating discriminative joint feature representation. An off-the-shelf classifier was trained on the transformed features which then produced the final class label.

For the *unsupervised* classification of lung nodules and IPMN, we proposed a learning approach based on label proportions. We first obtained an initial set of labels with clustering, followed by calculating label proportions using the cluster assignments. Finally, a proportion-SVM classifier was trained to generate final class labels. This work is significant to address the challenge of obtaining labeled medical images required to develop a machine learning model.

7.3 Future Directions

As deep learning based approaches permeate through almost all visual tasks, they can be applied for the quantification of white fat. In contrast to our proposed unsupervised strategy for SAT-VAT segmentation, the deep learning approaches, however, may require a large number of annotated examples to train a segmentation architecture. To address this issue, data augmentation can be performed either by image transformation or by adding noise. An alternative approach can be to use Generative Adversarial Networks to produce synthetic examples. We have started work in this direction by using GANs to generate realistic samples which were then shown to radiologists for qualitative evaluation [126]. These deep learning based approaches are expected to be more accurate and would not require an explicit setting of observational heuristics.

In future, segmentation of brown fat can be performed with a joint PET-CT CNN framework. Rather than explicitly modeling the asymmetry between PET and CT for the segmentation of

brown fat, these asymmetry parameters can be learned from the data. With sufficient examples to learn the BAT-WAT distinction, the need to have a false positive removal can also be circumvented. As MRI is becoming more ubiquitous, it can be used to study white and brown adipose tissues and existing CAD approaches can be modified for MRI scans.

Risk stratification of lung nodules can be extended by training a 3D CNN end-to-end. Moreover, a single architecture can be employed to perform joint detection, segmentation, and diagnosis of lung nodules. A more pertinent clinical application would be to use sequential CT scans for automatic lung cancer diagnosis. A possible research direction is to use recurrent neural networks and LSTMs to perform prediction on sequential data. Incorporation of other imaging modalities such as PET can also be useful especially to compute the metabolic tumor volume.

The automatic diagnosis of IPMN is a new yet very important problem for medical imaging researchers. The detection and segmentation of pancreas can be deemed useful to make a better prediction about the presence of IPMN and cysts. Due to its anatomy, pancreas is a challenging organ to segment, particularly in MRI images. To address this challenge, other imaging modalities such as CT can be utilized for joint segmentation and diagnosis of pancreatic cysts and IPMN. Visualization of activation maps can be quite useful for the clinicians to identify new imaging biomarkers that can be employed for diagnosis in future.

There is a lot of potential for research in developing unsupervised approaches for medical imaging applications. Medical imaging has unique challenges associated with the scarcity of labeled examples. Moreover, unless corroborated by biopsy, there may exist a large variability in labeling from different radiologists. Although transfer learning has helped to address the lack of annotated examples, the performance is limited due to large differences in domains. It is comparatively easier to obtain scan level labels than slice level labels. In this regard, weakly supervised approaches such as multiple instance learning (MIL) can be of great value. Active learning can be another solution

to alleviate the difficulty in labeling. Deep learning for joint feature learning and clustering can be employed to obtain data-driven clusters [127]. Reinforcement learning is also quite promising for medical imaging applications. In future, these research directions can be pursued to address unique medical imaging challenges and to have improved diagnostic decisions in clinics.

LIST OF REFERENCES

[1] Delbeke, D., Segall, G.M.: Status of and trends in nuclear medicine in the United States. Journal of Nuclear Medicine 52(Supplement 2), 24S–28S (2011)

[2] Arora, R.: The training and practice of radiology in India: current trends. Quantitative imaging in medicine and surgery 4(6), 449 (2014)

[3] Digital X-Ray on-the-go in Kenya. `https://radiology.ucsf.edu/blog/digital-x-ray-go-kenya`

[4] McDonald, R.J., Schwartz, K.M., Eckel, L.J., Diehn, F.E., Hunt, C.H., Bartholmai, B.J., Erickson, B.J., Kallmes, D.F.: The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. Academic radiology 22(9), 1191–1198 (2015)

[5] Lee, C.S., Nagy, P.G., Weaver, S.J., Newman-Toker, D.E.: Cognitive and system factors contributing to diagnostic errors in radiology. American Journal of Roentgenology 201(3), 611–617 (2013)

[6] Leape, L.L., Berwick, D.M., Bates, D.W.: Counting deaths due to medical errorsreply. JAMA 288(19), 2405–2405 (2002)

[7] Gunderman, R.B.: Biases in radiologic reasoning. American Journal of Roentgenology 192(3), 561–564 (2009)

[8] Preston, S.H.: Deadweight?– The influence of obesity on longevity. New England Journal of Medicine 352(11), 1135–1137 (2005)

[9] Cefalu, W.T., Werbel, S., Bell-Farrow, A.D., Terry, J.G., Wang, Z.O., Opara, E.C., Morgan, T., Hinson, W.H., Crouse, J.R.: Insulin resistance and fat patterning with aging: Relationship to metabolic risk factors for cardiovascular disease. Metabolism 47(4), 401–408 (1998)

[10] Gastaldelli, A., Miyazaki, Y., Pettiti, M., Matsuda, M., Mahankali, S., Santini, E., DeFronzo, R.A., Ferrannini, E.: Metabolic effects of visceral fat accumulation in type 2 diabetes. The Journal of Clinical Endocrinology & Metabolism 87(11), 5098–5103 (2002)

[11] Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., Mullany, E.C., Biryukov, S., Abbafati, C., Abera, S.F., et al.: Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the Global Burden of Disease Study 2013. The Lancet 384(9945), 766–781 (2014)

[12] National Center for Health Statistics and Centers for Disease Control and Prevention: Health, United States, 2015, with special feature on Racial and Ethnic health disparities (2015)

[13] Von Hafe, P., Pina, F., Pérez, A., Tavares, M., Barros, H.: Visceral Fat Accumulation as a Risk Factor for Prostate Cancer. Obesity 12(12), 1930 (2004)

[14] Moon, H.G., Ju, Y.T., Jeong, C.Y., Jung, E.J., Lee, Y.J., Hong, S.C., Ha, W.S., Park, S.T., Choi, S.K.: Visceral Obesity may affect Oncologic Outcome in patients with Colorectal Cancer. Annals of Surgical Oncology 15(7), 1918–1922 (2008)

[15] Britton, K.A., Massaro, J.M., Murabito, J.M., Kreger, B.E., Hoffmann, U., Fox, C.S.: Body Fat Distribution, Incident Cardiovascular Disease, Cancer, and All-Cause Mortality. Journal of the American College of Cardiology 62(10), 921–925 (2013)

[16] Speliotes, E.K., Massaro, J.M., Hoffmann, U., Vasan, R.S., Meigs, J.B., Sahani, D.V., Hirschhorn, J.N., O'Donnell, C.J., Fox, C.S.: Fatty liver is associated with dyslipidemia

and dysglycemia independent of visceral fat: the Framingham Heart Study. Hepatology 51(6), 1979–1987 (2010)

[17] Kuk, J.L., Katzmarzyk, P.T., Nichaman, M.Z., Church, T.S., Blair, S.N., Ross, R.: Visceral fat is an independent predictor of all-cause mortality in men. Obesity 14(2), 336–341 (2006)

[18] Tong, Y., Udupa, J.K., Torigian, D.A.: Optimization of abdominal fat quantification on CT imaging through use of standardized anatomic space: A novel approach. Medical physics 41(6), 063501 (2014)

[19] Cypess, A.M., Lehman, S., Williams, G., Tal, I., Rodman, D., Goldfine, A.B., Kuo, F.C., Palmer, E.L., Tseng, Y.H., Doria, A., Kolodny, G.M., Kahn, C.R.: Identification and Importance of Brown Adipose Tissue in Adult Humans. New England Journal of Medicine 360(15), 1509–1517 (2009)

[20] Stewart, B., Wild, C.P., et al.: World Cancer Report 2014. World (2016)

[21] Howlader, N., Noone, A., Krapcho, M., et al., G.: SEER Cancer Statistics Review, 1975-2011. National Cancer Institute. Bethesda, MD (2014)

[22] van Beek, E.J., Mirsadraee, S., Murchison, J.T.: Lung cancer screening: Computed tomography or chest radiographs? World journal of radiology 7(8), 189 (2015)

[23] Armato III, S., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical Physics 38(2), 915–931 (2011)

[24] Society, A.C.: Cancer Facts & Figures. American Cancer Society (2016)

[25] Shi, C., Hruban, R.H.: Intraductal papillary mucinous neoplasm. Human pathology 43(1), 1–16 (2012)

[26] Sadot, E., Basturk, O., Klimstra, D.S., Gönen, M., Anna, L., Do, R.K.G., et al.: Tumor-associated neutrophils and malignant progression in intraductal papillary mucinous neoplasms: an opportunity for identification of high-risk disease. Annals of surgery 262(6), 1102 (2015)

[27] Matthaei, H., Schulick, R.D., Hruban, R.H., Maitra, A.: Cystic precursors to invasive pancreatic cancer. Nature Reviews Gastroenterology and Hepatology 8(3), 141–150 (2011)

[28] Tanaka, M., Fernández-del Castillo, C., Adsay, V., Chari, S., Falconi, M., Jang, J.Y., Kimura, W., Levy, P., Pitman, M.B., Schmidt, C.M., et al.: International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology 12(3), 183–197 (2012)

[29] Sivic, J., Russell, B.C., Efros, A.A., Zisserman, A., Freeman, W.T.: Discovering objects and their location in images. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. vol. 1, pp. 370–377. IEEE (2005)

[30] Alayrac, J.B., Bojanowski, P., Agrawal, N., Sivic, J., Laptev, I., Lacoste-Julien, S.: Unsupervised learning from narrated instruction videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4575–4583 (2016)

[31] Hussein, S., Green, A., Watane, A., Reiter, D., Chen, X., Papadakis, G.Z., Wood, B., Cypess, A., Osman, M., Bagci, U.: Automatic segmentation and quantification of white and brown adipose tissues from PET/CT scans. IEEE Transactions on Medical Imaging 36(3), 734–744 (2017)

[32] Green, A.L., Bagci, U., Hussein, S., Kelly, P.V., Muzaffar, R., Neuschwander-Tetri, B.A., Osman, M.M.: Brown adipose tissue detected by PET/CT imaging is associated with less central obesity. Nuclear Medicine Communications 38(7), 629–635 (2017)

[33] Hussein, S., Gillies, R., Cao, K., Song, Q., Bagci, U.: Tumornet: Lung nodule characterization using multi-view convolutional neural network with gaussian process. In: Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. pp. 1007–1010. IEEE (2017)

[34] Hussein, S., Cao, K., Song, Q., Bagci, U.: Risk stratification of lung nodules using 3D CNN-based multi-task learning. In: International Conference on Information Processing in Medical Imaging. pp. 249–260. Springer (2017)

[35] Hussein, S., Kandel, P., Corral, J.E., Bolan, C.W., Wallace, M.B., Bagci, U.: Deep multimodal classification of intraductal papillary mucinous neoplasms (ipmn) with canonical correlation analysis. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. pp. 800–804. IEEE (2018)

[36] Zhao, B., Colville, J., Kalaigian, J., Curran, S., Jiang, L., Kijewski, P., Schwartz, L.H.: Automated Quantification of Body Fat Distribution on Volumetric Computed Tomography. Journal of Computer Assisted Tomography 30(5), 777–783 (2006)

[37] Romero, D., Ramirez, J.C., Mármol, A.: Quanification of subcutaneous and visceral adipose tissue using CT. In: IEEE International Workshop on Medical Measurement and Applications, 2006. MeMea 2006. pp. 128–133. IEEE (2006)

[38] Pednekar, A., Bandekar, A.N., Kakadiaris, I.A., Naghavi, M.: Automatic Segmentation of Abdominal Fat from CT Data. In: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)-Volume 01. pp. 308–315. IEEE Computer Society (2005)

[39] Mensink, S.D., Spliethoff, J.W., Belder, R., Klaase, J.M., Bezooijen, R., Slump, C.H.: Development of automated quantification of visceral and subcutaneous adipose tissue volumes

from abdominal CT scans. In: SPIE Medical Imaging. pp. 79632Q–79632Q. International Society for Optics and Photonics (2011)

[40] Kim, Y.J., Lee, S.H., Kim, T.Y., Park, J.Y., Choi, S.H., Kim, K.G.: Body fat assessment method using CT images with separation mask algorithm. Journal of Digital Imaging 26(2), 155–162 (2013)

[41] Chung, H., Cobzas, D., Birdsell, L., Lieffers, J., Baracos, V.: Automated segmentation of muscle and adipose tissue on CT images for human body composition analysis. In: SPIE Medical Imaging. pp. 72610K–72610K. International Society for Optics and Photonics (2009)

[42] Kim, Y.J., Park, J.W., Kim, J.W., Park, C.S., Gonzalez, J.P.S., Lee, S.H., Kim, K.G., Oh, J.H.: Computerized Automated Quantification of Subcutaneous and Visceral Adipose Tissue From Computed Tomography Scans: Development and Validation Study. Journal of Medical Internet Research; Medical Informatics 4(1) (2016)

[43] Muzik, O., Mangner, T.J., Leonard, W.R., Kumar, A., Janisse, J., Granneman, J.G.: ^{15}O PET Measurement of Blood Flow and Oxygen Consumption in Cold-Activated Human Brown Fat. Journal of Nuclear Medicine 54(4), 523–531 (2013)

[44] Cohade, C., Osman, M., Pannu, H., Wahl, R.: Uptake in supraclavicular area fat ("USA-Fat"): Description on ^{18}F-FDG PET/CT. Journal of Nuclear Medicine 44(2), 170–176 (2003)

[45] Baba, S., Jacene, H.A., Engles, J.M., Honda, H., Wahl, R.L.: CT Hounsfield Units of Brown Adipose Tissue Increase with Activation: Preclinical and Clinical Studies. Journal of Nuclear Medicine 51(2), 246–250 (2010)

[46] Gilsanz, V., Chung, S.A., Jackson, H., Dorey, F.J., Hu, H.H.: Functional Brown Adipose Tissue is Related to Muscle Volume in Children and Adolescents. The Journal of Pediatrics pp. 722–726 (2011)

[47] Uchiyama, Y., Katsuragawa, S., Abe, H., Shiraishi, J., Li, F., Li, Q., Zhang, C.T., Suzuki, K., Doi, K.: Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. Medical Physics 30(9), 2440–2454 (2003)

[48] Farag, A., Ali, A., Graham, J., Farag, A., Elshazly, S., Falk, R.: Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. In: 2011 IEEE International Symposium on Biomedical Imaging: from nano to macro. pp. 169–172. IEEE (2011)

[49] Sorensen, L., Shaker, S.B., De Bruijne, M.: Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Transactions on Medical Imaging 29(2), 559–569 (2010)

[50] Song, Y., Cai, W., Zhou, Y., Feng, D.D.: Feature-based image patch approximation for lung tissue classification. IEEE Transactions on Medical Imaging 32(4), 797–808 (2013)

[51] Orozco, H.M., Villegas, O.O.V., Sánchez, V.G.C., Domínguez, H.d.J.O., Alfaro, M.d.J.N.: Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. Biomedical Engineering Online 14(1), 1 (2015)

[52] Ma, J., Wang, Q., Ren, Y., Hu, H., Zhao, J.: Automatic lung nodule classification with radiomics approach. In: SPIE Medical Imaging. pp. 978906–978906 (2016)

[53] Kumar, D., Wong, A., Clausi, D.A.: Lung nodule classification using deep features in CT images. In: Computer and Robot Vision (CRV), 2015 12th Conference on. pp. 133–138. IEEE (2015)

[54] Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., Mollura, D.J.: Characterization of lung nodule malignancy using hybrid shape and appearance features. In: MICCAI (2016)

[55] El-Baz, A., Nitzken, M., Khalifa, F., Elnakib, A., Gimelfarb, G., Falk, R., El-Ghar, M.A.: 3D shape analysis for early diagnosis of malignant lung nodules. In: Biennial International Conference on Information Processing in Medical Imaging. pp. 772–783. Springer (2011)

[56] Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H., Liang, Z.: Texture feature analysis for computer-aided diagnosis on pulmonary nodules. Journal of Digital Imaging 28(1), 99–115 (2015)

[57] Way, T.W., Hadjiiski, L.M., Sahiner, B., Chan, H.P., Cascade, P.N., Kazerooni, E.A., Bogot, N., Zhou, C.: Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. Medical Physics 33(7), 2323–2337 (2006)

[58] Lee, M., Boroczky, L., Sungur-Stasik, K., Cann, A., Borczuk, A., Kawut, S., Powell, C.: Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction. Artificial intelligence in medicine 50(1), 43–53 (2010)

[59] Chen, S., Ni, D., Qin, J., Lei, B., Wang, T., Cheng, J.Z.: Bridging Computational Features Toward Multiple Semantic Features with Multi-task Regression: A Study of CT Pulmonary Nodules. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 53–60. Springer (2016)

[60] Furuya, K., Murayama, S., Soeda, H., Murakami, J., Ichinose, Y., Yauuchi, H., Katsuda, Y., Koga, M., Masuda, K.: New classification of small pulmonary nodules by margin characteristics on highresolution CT. Acta Radiologica 40(5), 496–504 (1999)

[61] Hanania, A.N., Bantis, L.E., Feng, Z., Wang, H., Tamm, E.P., Katz, M.H., Maitra, A., Koay, E.J.: Quantitative imaging to evaluate malignant potential of IPMNs. Oncotarget 7(52), 85776 (2016)

[62] Gazit, L., Chakraborty, J., Attiyeh, M., Langdon-Embry, L., Allen, P.J., Do, R.K., Simpson, A.L.: Quantification of CT Images for the Classification of High-and Low-Risk Pancreatic Cysts. In: SPIE Medical Imaging. pp. 101340X–101340X. International Society for Optics and Photonics (2017)

[63] Quan, C., Wang, M., Ren, F.: An unsupervised text mining method for relation extraction from biomedical literature. PloS one 9(7), e102039 (2014)

[64] Kamper, H., Jansen, A., Goldwater, S.: Fully unsupervised small-vocabulary speech recognition using a segmental Bayesian model. In: Interspeech (2015)

[65] Shin, H.C., Orton, M.R., Collins, D.J., Doran, S.J., Leach, M.O.: Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(8), 1930–1943 (2013)

[66] Vaidhya, K., Thirunavukkarasu, S., Alex, V., Krishnamurthi, G.: Multi-modal brain tumor segmentation using stacked denoising autoencoders. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 181–194. Springer (2015)

[67] Sivakumar, S., Chandrasekar, C.: Lung nodule segmentation through unsupervised clustering models. Procedia engineering 38, 3064–3073 (2012)

[68] Zhang, F., Song, Y., Cai, W., Zhou, Y., Fulham, M., Eberl, S., Shan, S., Feng, D.: A ranking-based lung nodule image classification method using unlabeled image knowledge.

In: Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. pp. 1356–1359. IEEE (2014)

[69] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105 (2012)

[70] Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 35(5), 1285–1298 (2016)

[71] Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: Delving deep into convolutional nets. In: British Machine Vision Conference (2014)

[72] Hinton, G.E.: Products of experts. In: Ninth International Conference on Artificial Neural Networks, 1999 (Conf. Publ. No. 470). vol. 1, pp. 1–6. IET (1999)

[73] Yoshizumi, T., Nakamura, T., Yamane, M., Waliul Islam, A.H.M., Menju, M., Yamasaki, K., Arai, T., Kotani, K., Funahashi, T., Yamashita, S., et al.: Abdominal Fat: Standardized Technique for Measurement at CT. Radiology 211(1), 283–286 (1999)

[74] Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49(4), 764–766 (2013)

[75] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 886–893. IEEE (2005)

[76] Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research 9(Nov), 2579–2605 (2008)

[77] Kriegel, H.P., Kröger, P., Schubert, E., Zimek, A.: LoOP: Local Outlier Probabilities. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management. pp. 1649–1652. ACM (2009)

[78] Boykov, Y., Veksler, O., Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), 1222–1239 (2001)

[79] Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23(7), 903–921 (2004)

[80] Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Transactions on Medical Imaging 29(10), 1714–1729 (2010)

[81] Udupa, J.K., Leblanc, V.R., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L.M., Hirsch, B.E., Woodburn, J.: A framework for evaluating image segmentation algorithms. Computerized Medical Imaging and Graphics 30(2), 75–87 (2006)

[82] Kohlberger, T., Singh, V., Alvino, C., Bahlmann, C., Grady, L.: Evaluating segmentation error without ground truth. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528–536. Springer (2012)

[83] Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: SIFT Flow: Dense correspondence across different scenes. In: European Conference on Computer Vision, pp. 28–42 (2008)

[84] Jain, M., Van Gemert, J., Jégou, H., Bouthemy, P., Snoek, C.G.: Action localization with tubelets from motion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 740–747 (2014)

[85] Udupa, J.K., Odhner, D., Zhao, L., Tong, Y., Matsumoto, M.M., Ciesielski, K.C., Falcao, A.X., Vaideeswaran, P., Ciesielski, V., Saboury, B., et al.: Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. Medical Image Analysis 18(5), 752–771 (2014)

[86] Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981)

[87] Sugiyama, M., Borgwardt, K.: Rapid distance-based outlier detection via sampling. In: Advances in Neural Information Processing Systems (NIPS). pp. 467–475 (2013)

[88] Bagci, U., Udupa, J.K., Mendhiratta, N., Foster, B., Xu, Z., Yao, J., Chen, X., Mollura, D.J.: Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Medical Image Analysis 17(8), 929–945 (2013)

[89] Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)

[90] Browne, J., De Pierro, A.: A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. IEEE Transactions on Medical Imaging 15(5), 687–699 (1996)

[91] Ono, K., Hiraoka, T., Ono, A., Komatsu, E., Shigenaga, T., Takaki, H., Maeda, T., Ogusu, H., Yoshida, S., Fukushima, K., et al.: Low-dose CT scan screening for lung cancer: Com-

parison of images and radiation doses between low-dose CT and follow-up standard diagnostic CT. SpringerPlus 2(1), 1 (2013)

[92] Orenstein, B.W.: Reducing PET Dose. Radiology Today 17(1), 22 (2015)

[93] Akin, E.A., Torigian, D.A.: Considerations regarding radiation exposure in performing FDG-PET-CT. Image Wisely (2012)

[94] Prakash, K.B., Srour, H., Velan, S.S., Chuang, K.H.: A method for the automatic segmentation of Brown Adipose Tissue. Magnetic Resonance Materials in Physics, Biology and Medicine 29(2), 287–299 (2016)

[95] Flynn, A., Li, Q., Panagia, M., Abdelbaky, A., MacNabb, M., Samir, A., Cypess, A.M., Weyman, A.E., Tawakol, A., Scherrer-Crosbie, M.: Contrast-Enhanced Ultrasound: A Novel Noninvasive, Nonionizing Method for the Detection of Brown Adipose Tissue in Humans. Journal of the American Society of Echocardiography 28(10), 1247–1254 (2015)

[96] Xu, Z., Bagci, U., Seidel, J., Thomasson, D., Solomon, J., Mollura, D.J.: Segmentation based denoising of PET images: An iterative approach via regional means and affinity propagation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 698–705. Springer (2014)

[97] Xu, Z., Bagci, U., Gao, M., Mollura, D.J.: Highly precise partial volume correction for PET images: An iterative approach via shape consistency. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). pp. 1196–1199. IEEE (2015)

[98] Wang, H., Udupa, J.K., Odhner, D., Tong, Y., Zhao, L., Torigian, D.A.: Automatic anatomy recognition in whole-body PET/CT images. Medical Physics 43(1), 613–629 (2016)

[99] Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: Learning spatiotemporal features with 3D convolutional networks. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 4489–4497. IEEE (2015)

[100] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 1725–1732 (2014)

[101] Evgeniou, T., Pontil, M.: Regularized multi–task learning. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge discovery and Data mining. pp. 109–117. ACM (2004)

[102] Zhou, J., Chen, J., Ye, J.: MALSAR: Multi-task learning via structural regularization (2012)

[103] Nesterov, Y.: Introductory lectures on convex optimization: A basic course, vol. 87. Springer Science & Business Media (2013)

[104] Parikh, N., Boyd, S.P., et al.: Proximal algorithms

[105] Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42(3), 145–175 (2001)

[106] Wang, Y.X.J., Gong, J.S., Suzuki, K., Morcos, S.K.: Evidence based imaging strategies for solitary pulmonary nodule. Journal of thoracic disease 6(7), 872–887 (2014)

[107] Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Fully automatic face normalization and single sample face recognition in unconstrained environments. Expert Systems with Applications 47, 23–34 (2016)

[108] Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging 29(6), 1310–1320 (2010)

[109] Johnson, H.J., McCormick, M., Ibáñez, L., Consortium, T.I.S.: The ITK Software Guide. Kitware, Inc., third edn. (2013)

[110] He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. pp. 1322–1328. IEEE (2008)

[111] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)

[112] Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., Matrisian, L.M.: Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer research 74(11), 2913–2921 (2014)

[113] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680 (2014)

[114] Yu, F., Liu, D., Kumar, S., Tony, J., Chang, S.F.: $\propto$SVM for Learning with Label Proportions. In: Proceedings of The 30th International Conference on Machine Learning. pp. 504–512 (2013)

[115] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)

[116] Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery.

In: International Conference on Information Processing in Medical Imaging. pp. 146–157. Springer (2017)

[117] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802 (2016)

[118] Yeh, R., Chen, C., Lim, T.Y., Hasegawa-Johnson, M., Do, M.N.: Semantic image inpainting with perceptual and contextual losses. arXiv preprint arXiv:1607.07539 (2016)

[119] Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. In: Advances In Neural Information Processing Systems. pp. 613–621 (2016)

[120] Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., Zhavoronkov, A.: The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget 8(7), 10883 (2017)

[121] Odena, A.: Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv:1606.01583 (2016)

[122] Nguyen, A., Yosinski, J., Bengio, Y., Dosovitskiy, A., Clune, J.: Plug & play generative networks: Conditional iterative generation of images in latent space. arXiv preprint arXiv:1612.00005 (2016)

[123] Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. In: 33rd International Conference on Machine Learning (ICML 2016) International Conference on Machine Learning (2016)

[124] Springenberg, J.T.: Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv preprint arXiv:1511.06390 (2015)

[125] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2172–2180 (2016)

[126] Chuquicusma, M.J., Hussein, S., Burt, J., Bagci, U.: How to Fool Radiologists with Generative Adversarial Networks? A Visual Turing Test for Lung Cancer Diagnosis. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE (2018)

[127] Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International conference on machine learning. pp. 478–487 (2016)