THE UNIVERSITY OF CHICAGO

FUNCTIONAL, VOLUMETRIC, AND TEXTURAL ANALYSIS OF MALIGNANT PLEURAL MESOTHELIOMA USING COMPUTED TOMOGRAPHY AND DEEP CONVOLUTIONAL NEURAL NETWORKS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON MEDICAL PHYSICS

BY
EYJÓLFUR GUÐMUNDSSON

CHICAGO, ILLINOIS
AUGUST 2019

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to start by thanking my advisor, Samuel Armato, Ph.D., for his guidance, mentorship and support throughout my dissertation research. Sam is an exceptionally supportive advisor and a researcher of rigor and integrity. I am happy to have had the opportunity to learn from him in the close capacity of thesis advisor and student. I would like to acknowledge my dissertation committee members, Maryellen Giger, Ph.D., Christopher Straus, M.D., Hedy Kindler, M.D., and Aliya Husain, M.D., for their valuable insights, which helped improve the technical quality and clinical focus of my work. In particular, I would like to thank Chris for the generous allocation of his time and expertise to teach me and assist me in my research work at this institution.

I would like to thank my fellow lab members, Joe Foy, Adam Starkey, Fawwaz Qayyum, Roger Engelmann, and Feng Li, M.D., Ph.D., for their helpful insights and support throughout my dissertation research. In particular, Feng Li's assistance in the collection and curation of the data sets of the third and fourth specific aims of my dissertation research was invaluable; her work was essential in helping me achieve what I had set out to do in these two projects. I would also like to thank my collaborators in the Department of Pathology, Dr. Husain and David Chapel, M.D., for their collaboration on the fourth aim of my dissertation research, and the researchers and clinical staff of the mesothelioma program at the University of Chicago, headed by Dr. Kindler, for their assistance and support throughout my research on the image-based analysis of this disease.

I would like to recognize the support and efforts of the knowledgable faculty members of the Graduate Program in Medical Physics toward my education and research during my time here. I would like to thank all of my fellow students in the program, past and present, for their friendship and support – in particular, my classmates, Natasha Antropova, Greg Anthony, and Corey Smith, as well as Neville Eclov and Fred Pineda, and many others. I would like to thank the administrative staff of the graduate program and the division for their assistance throughout my time at the University. My gratitude to my friends from

other departments of the University and from around Chicago, including João Caldeira and Mathilde Gerbelli-Gauthier, and to my colleagues in Graduate Students United, for their friendship and support during my time in Chicago.

The support and encouragement of my family and friends has been crucial for the successful completion of my dissertation research at the University of Chicago. I was happy to host those of my family and friends who got the chance to visit me in Chicago, some of whom visited me on multiple occasions during my five years here. To my parents and both of my brothers, thank you for your support in this undertaking – one of my most memorable trips was with my brother Snæbjörn, who traveled with me thousands of miles across the country, from Seattle to Chicago, in search of incredible scenery and interesting places. I would like to acknowledge the support of my former colleagues at the medical engineering company Raförninn in Reykjavik, in particular Smári Kristinsson and Hildur Ólafsdóttir; without the introduction to the field of medical physics and Smári's encouragement during my summer research internships and subsequent full-time work there, it's quite likely that I would not have pursued a graduate degree in medical physics. To my partner Eszter Czibor, who is herself a brilliant researcher and an incredibly warm, generous, and thoughtful individual, thank you for all your support and encouragement, and the many insightful and enlightening conversations we have had for the past two years. I am looking forward to our next steps together.

To those who have been named here and all others who have assisted me in this endeavor, my gratitude for your support.

ABSTRACT

Malignant pleural mesothelioma (mesothelioma) is a malignancy of the pleura, a membrane lining the outside of the lungs and the inside of the chest wall. This thoracic cancer is primarily caused by asbestos exposure and carries a poor prognosis. Computed tomography (CT) is the main imaging modality used for the diagnosis and treatment evaluation of mesothelioma in the United States. Mesothelioma is typically aspherical and presents with different morphology and growth patterns from most other cancers. These aspects of the disease, combined with recent advances in personalized therapies and molecularly targeted therapies, have motivated investigations into alternative and potentially more sensitive measures of mesothelioma treatment response. Image-based measurement of mesothelioma tumor volume has seen interest from researchers in the staging of this disease and as a potential marker for patient prognosis and tumor response to treatment. The histologic subtype of mesothelioma is the most significant prognostic factor for this disease.

The first study presented in this work investigated the use of functional CT imaging for the assessment of mesothelioma response to treatment. This prospective study showed differing trends in hemodynamic change between scans for patients on treatment and patients on observation, which suggests a potential role of this imaging method for the assessment of tumor response to treatment. A future study of a larger patient population on more coherent treatment regimens is needed to confirm the results of this pilot study. In another investigation of this work, deep convolutional neural networks (CNNs) were implemented for the automated segmentation of mesothelioma tumor on CT scans, showing a significantly improved segmentation performance when compared with a previously published non-deep learning-based method for the automated segmentation of this disease. This study demonstrated the suitability and state-of-the-art performance of deep CNNs for the task of mesothelioma tumor segmentation on CT scans. Next, a study investigated the performance of a deep CNN-based tumor segmentation method on CT scans that exhibited both mesothelioma tumor and pleural effusions, a common presentation on radiologic scans of mesothelioma patients.

This study showed significantly improved tumor segmentation performance when compared with the previous deep CNN-based method on scans of mesothelioma patients exhibiting tumor and pleural effusions. The last study presented in this work explored the differentiation of histologic subtypes of mesothelioma on CT scans using texture analysis through a retrospective investigation. CT scans of patients with pathologically confirmed epithelioid and "non-epithelioid" mesothelioma (i.e., biphasic and sarcomatoid tumors), acquired prior to talc pleurodesis, were included in the final analysis. The database of this study included in-house scans and outside scans acquired on a variety of CT scanners, and data heterogeneity precluded the robust classification of tumors as epithelioid and non-epithelioid; however, the method warrants further investigation as more data become available.

This dissertation investigated methods for the quantitative image-based analysis of mesothelioma tumor through standard contrast-enhanced and functional imaging and through the acquisition of image-based tumor volume using advanced deep learning-based algorithms. In this work, the novel investigation of deep CNNs for the segmentation of mesothelioma on CT scans demonstrated the state-of-the-art performance of this technique for the automated acquisition of mesothelioma tumor volumetrics, the first pilot study dedicated to the investigation of DCE-CT for mesothelioma treatment response assessment indicated that this method should be explored in a larger patient cohort, and texture feature analysis was explored for the first time for the differentiation of mesothelioma tumor subtype in a retrospective study. The research presented in this work will hopefully advance the methods available to researchers and clinicians for the quantitative analysis of mesothelioma in the image-based management of mesothelioma patients.

CHAPTER 1

INTRODUCTION

1.1 Malignant Pleural Mesothelioma

Malignant mesothelioma is an aggressive cancer diagnosed annually in approximately 3,000 patients in the United States that can arise from the mesothelial surfaces of the pleura, peritoneum, pericardium and tunica vaginalis [2, 3]. Approximately 85% of cases involve the pleura, the membrane that forms the outer lining of the lungs (the visceral pleura) and the inner lining of the thoracic cavity (the parietal pleura) [3]. For the purposes of this work, "mesothelioma" will refer to malignant pleural mesothelioma.

1.1.1 Epidemiology

Malignant pleural mesothelioma is most commonly caused by exposure to asbestos; at least 70% of mesothelioma patients have a definite history of prior asbestos exposure [3, 4]. Other causes of mesothelioma include prior radiation exposure and a genetic predisposition to the disease in individuals exposed to the mineral erionite [2]. "Asbestos" is a collective term used for a group of naturally occurring fibrous minerals, which have seen various applications in human history owing to their heat-resistant properties, tensile strength and fibrous appearance [5]. This group of minerals was first described as a cause of mesothelioma in a group of miners in South Africa in 1960 [6, 7]. Demand for asbestos increased with industrialization, and the latter half of the 20th century saw an epidemic of mesothelioma, particularly in developed countries as a result of asbestos production and use in industry [4]. As the associated health hazards became more widely known, the US saw increased exposure controls of the mineral in the 1970s; however, due to the long latency period of 20–60 years between first exposure to asbestos and the clinical presentation of mesothelioma, the incidence of the disease kept increasing after exposure limits were introduced [4, 7, 8]. The occurrence of mesothelioma is 2 per million per year in women and 10–30 per million per year in men; due

to the occupational nature of this disease, it occurs more commonly in men [3, 7, 8]. The increase in mesothelioma incidence has leveled off only in countries that established asbestos control measures in the 1970s, and in countries that continue to produce and/or use asbestos, a rise in incidence of this disease could be expected [3, 7].

1.1.2 Diagnosis and Prognosis

Patients with mesothelioma most commonly present with dyspnea and/or chest pain [2, 3]. Chest radiography is usually the first radiological examination performed and typically reveals a unilateral pleural effusion, which is a common cause of breathlessness in early stages of the disease [3, 9]. Radiographs are insensitive and nonspecific, however, and require follow-up examination by other imaging modalities (see Section 1.1.3). Mesothelioma is more commonly found in the right hemithorax, and fewer than 5% of patients present with disease on both sides of the chest [1, 3].

Tumor histologic subtype is the most significant prognostic factor for mesothelioma [3, 10]. Pleural fluid cytology has low sensitivity in the diagnosis of mesothelioma; therefore, tumor biopsy through thoracoscopy is the gold standard for the diagnosis of this disease [3, 10]. The main histologic subtypes of mesothelioma are epithelioid, sarcomatoid and biphasic (mixed subtype); non-epithelioid subtype signifies poorer patient prognosis [10]. Epithelioid tumors constitute 60–70% of cases, while approximately 10% of mesothelioma tumors are of the sarcomatoid subtype [3]. In the analysis of an international database of mesothelioma patients who had surgery as part of their care, patients with epithelioid mesothelioma were found to have a median survival of 19 months, patients with tumors of the biphasic subtype a median survival of 13 months and sarcomatoid patients a median survival of 8 months [11]. In addition to the differences in survival among the three main histologic subtypes of mesothelioma, recent research has indicated the prognostic value of a more detailed histologic subtyping and nuclear grading of epithelioid tumor as well as the amount of epithelioid differentiation in biphasic tumors [12, 13]. Non-tissue-based biomarkers in

mesothelioma are currently of limited sensitivity and specificity; soluble mesothelin-related protein has shown promise as a diagnostic biomarker but lacks sensitivity [14]. Research has suggested that loss of the BRCA1-associated protein (BAP1) is associated with better prognosis in mesothelioma [15, 16].

1.1.3 Imaging

As stated in Section 1.1.2, chest radiography is typically the first imaging examination in cases of mesothelioma, although further imaging is usually required [9]. Pleural fluid in one hemithorax is the most common initial manifestation of the disease on radiographs and is present in approximately 70% of patients [1, 4, 17].

Computed tomography (CT) is the primary imaging modality for the diagnosis, staging and evaluation of tumor response to treatment in mesothelioma. Pleural thickening and pleural effusion are the most common findings on chest CT scans of mesothelioma patients. The limitations of CT for the evaluation of mesothelioma include the evaluation of nodal involvement and the underestimation of tumor extent due to the limited contrast of tumor and adjacent normal soft tissues of the chest wall, mediastinum and diaphragm [18, 19]. CT has not been found to reliably distinguish among mesothelioma tumor subtypes, although ipsilateral volume loss has been found to be frequently associated with sarcomatoid and biphasic tumor [20].

^{18}F-fluorodeoxyglucose positron emission tomography/CT (FDG-PET/CT) allows for the combined evaluation of mesothelioma tumor morphology and metabolism. PET/CT can better distinguish enlargement of lymph nodes and metastatic disease when compared with other imaging modalities and has shown usefulness for the staging of mesothelioma [1, 21]. In some centers, FDG-PET/CT is used as a supplement to diagnosis and to confirm suspicion of pleural thickening as malignancy [4]. Furthermore, several studies have investigated the correlation between quantitative FDG-PET parameters and patient survival in mesothelioma [22, 23].

Magnetic resonance imaging (MRI) is not routinely used to evaluate mesothelioma in the US but may prove to be more sensitive than CT for detection of local invasion into adjacent structures, such as the mediastinum, chest wall and diaphragm [1, 24]. A single study found the apparent diffusion coefficient (ADC) values of epithelioid tumor, as measured using diffusion-weighted MRI (DWI-MRI), to be significantly higher than that of sarcomatoid tumor in a group of 50 mesothelioma patients [25], indicating the potential utility of imaging-based noninvasive diagnosis of histologic subtypes of this disease. Figure 1.1 shows examples of radiologic images of mesothelioma patients acquired with CT, MRI and PET-CT.

1.1.4 Staging

Staging is used to classify the anatomic extent, or stage, of cancer to assist in the determination of prognosis and appropriate treatment strategies for cancer patients. The most widely used staging system for mesothelioma was published in 1995 by the International Mesothelioma Interest Group (IMIG) and is based on the tumor, node, metastasis (TNM) staging system, in which the local extent of tumor, nodal involvement and metastatic disease determine the cancer stage [3, 26]. Revisions have recently been proposed for this staging system based on evidence acquired through a multinational database, which consists predominantly of data from patients who have undergone surgery for mesothelioma [27–30]. Current imaging techniques are limited in their ability to determine the TNM stage of mesothelioma tumors; therefore, definitive staging systems for this disease are surgically based [3, 9, 31]. No staging system for mesothelioma adequately predicts prognosis in patients who have not undergone tumor resection [2].

1.1.5 Treatment

Mesothelioma is an aggressive cancer for which no curative treatment currently exists. Treatments available to patients are cytotoxic chemotherapy, surgery and radiation therapy, and these options are often combined as part of a multimodality treatment strategy [2, 9, 32].

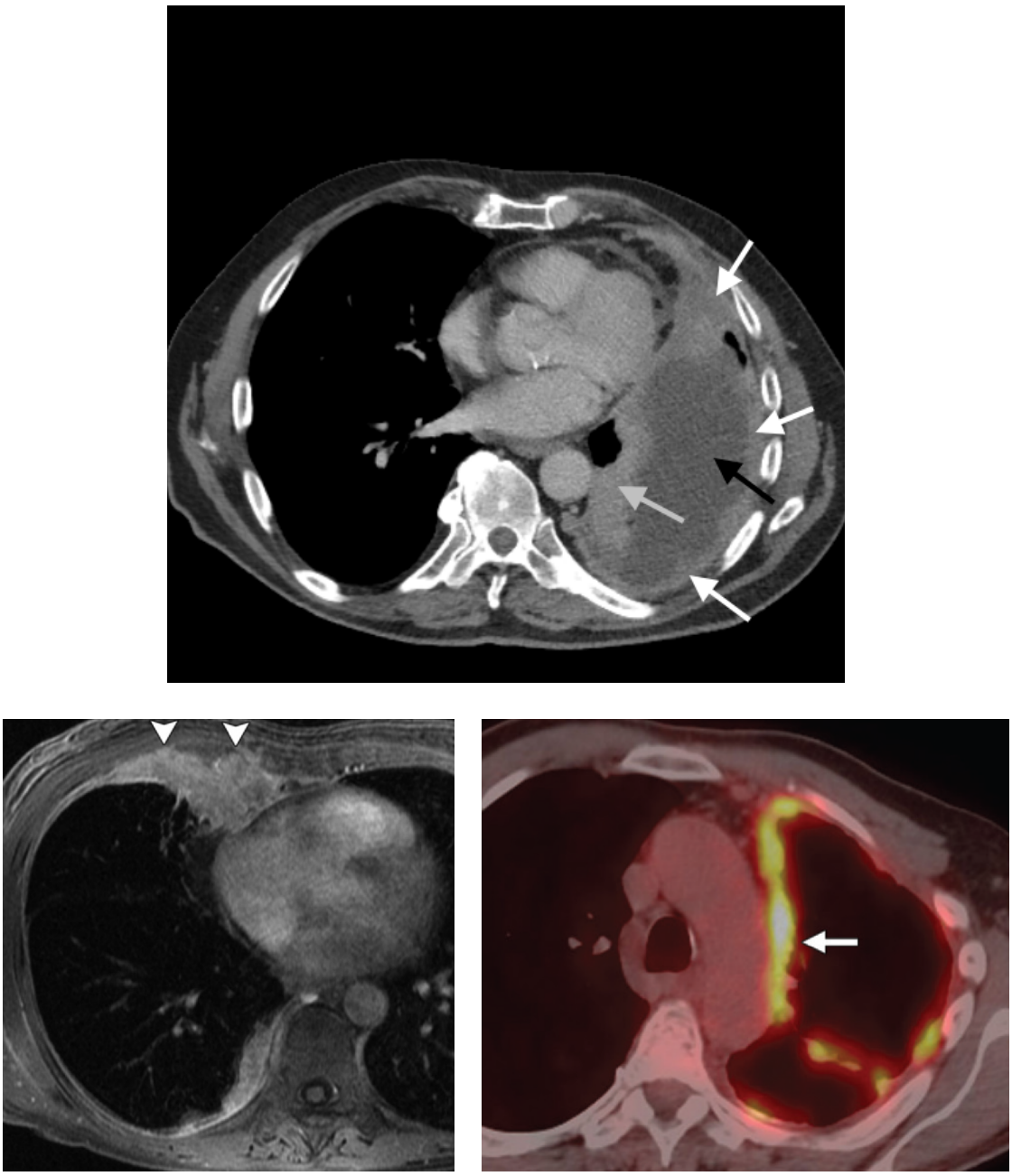

Figure 1.1: (Top) Axial CT section showing tumor (white arrow), pleural effusion (black arrow) and collapsed lung (gray arrow). (Bottom left) Axial slice from a contrast-enhanced T1-weighted MRI scan of a mesothelioma patient, showing chest wall invasion (white arrows; reprinted with permission from [1]). (Bottom right) Fused axial section from a PET/CT scan showing FDG-avid pleural thickening (white arrow; reprinted with permission from [1]).

Other treatment regimens that are under active investigation for mesothelioma include targeted therapies and anti-tumor immunotherapy [33].

Systemic chemotherapy is the only treatment option for the majority of mesothelioma patients [2]. The standard first-line chemotherapy in the treatment of mesothelioma has been a combination treatment of pemetrexed and cisplatin or pemetrexed and carboplatin since the results of a successful phase III clinical trial were published in 2003 [33, 34]. Based on this trial and control groups in other studies, the median overall survival for patients on such a treatment regimen is approximately 13–16 months [35, 36].

Approximately 20% of patients in the US undergo surgery for mesothelioma; however, clinicians lack consensus regarding the role of surgery in the treatment of this disease [3]. The two most common surgical procedures for curative intent of mesothelioma are extended pleurectomy/decortication (P/D) and extrapleural pneumonectomy (EPP); the aim of both procedures is macroscopic resection of tumor, with adjuvant therapy provided to treat any residual microscopic disease after surgery [2, 37]. EPP involves complete resection of the lung, pleura, pericardium and diaphragm of the affected side of the chest (leaving the pericardium and diaphragm intact in cases where these structures are not affected); extended P/D includes the resection of the pleura, pericardium and hemidiaphragm of the affected hemithorax, leaving the lung intact [38, 39]. The clinical outcomes of a randomized trial that was intended to assess EPP for the management of mesothelioma showed the procedure to potentially be harmful to patients; following this trial and other observational studies, this procedure has now been largely abandoned in favor of less radical surgical approaches, such as P/D [9, 37, 40].

The majority of mesothelioma patients experiences a pleural effusion at some point in their history with the disease; drainage of pleural fluid can alleviate the dyspnea that commonly occurs with pleural effusions [9, 16, 17]. Aspiration of pleural fluid can be achieved through a video-assisted thoracoscopy, while permanent fluid control may be realized through chemical pleurodesis. The aim of the latter procedure is fusion of the pleural space, accom-

plished through the introduction of a chemical irritant, most commonly talc [9, 41]. Talc pleurodesis has been found to cause pleural inflammation that may persist for up to six weeks following the procedure; furthermore, areas of high attenuation and of increased FDG uptake can appear on CT scans and PET-CT scans following talc pleurodesis [42–44]. A study on the outcomes of pleurodesis procedures in mesothelioma patients reported that less than 30% of patients who underwent pleurodesis achieved complete lifelong control of their effusion and approximately 30% required further drainage of pleural fluid [45].

Radiation therapy may be used for palliative treatment of mesothelioma or as an adjuvant therapy to surgery and chemotherapy in a multimodality setting [9]. The principal role of this modality in mesothelioma, due to the technical difficulties in the delivery of high doses of radiation to the pleura without causing damage to the underlying lung, is as an adjuvant and often focally targeted treatment following EPP [3].

Novel therapeutic strategies for mesothelioma include therapies that target specific growth factors and other biomarkers of the tumor and immunotherapies that aim to enhance the patient's immune response to the cancer. These types of therapies indicate a shift towards "personalized" treatments for mesothelioma, for which the preferred treatment regimen would depend on the genetic profile and biomarker expression of a patient's tumor; therefore, the success of targeted therapies generally relies on identification of patients most likely to benefit from the treatment [9].

Angiogenic growth factors are highly expressed in mesothelioma tumors, and the vascular endothelial growth factor (VEGF) has therefore been the focus of several studies of targeted therapies for this disease [33]. A large phase III clinical trial showed a significant increase in median overall survival with the anti-VEGF drug bevacizumab added to a treatment regimen of cisplatin and pemetrexed; however, other trials of antiangiogenic therapies have been less successful [33, 36]. Despite preclinical models for drugs that target mesothelioma biomarkers having shown early promise, targeted therapies for mesothelioma are hindered by the fact that a biomarker of high diagnostic sensitivity and specificity has not been found [3, 9].

Immune checkpoint inhibitors, which aim to reverse tumor suppression of immune response, have shown promise in both melanoma and non-small cell lung cancer and are being actively investigated for use in the mesothelioma setting [33]. Programmed cell death ligand-1 (PD-L1) is a molecule that inhibits the function of T cells of the immune system and is highly expressed by mesothelioma cells; several studies involving drugs that target this molecule are being carried out for treatment of mesothelioma [33].

1.1.6 Image-Based Assessment of Tumor Response to Treatment

Mesothelioma commonly grows as an irregular intermittent nodular pleural thickening or as a rind-like sheet of tumor around the lung and can extend into all aspects of the thoracic cavity [17]. This uncommon growth pattern sets mesothelioma apart from most other cancers and, combined with high case-to-case variability in presentation of the disease, complicates image-based assessment of tumor and of potential tumor response to treatment.

The current clinical measurement protocol for the evaluation of tumor response to treatment is the Modified Response Evaluation Criteria in Solid Tumors (RECIST) protocol [46, 47]. This measurement protocol was developed due to the incompatibility of mesothelioma tumor with standard nodular intrapulmonary malignancies, for which the longest-diameter measurement approach of the RECIST guidelines provide a reproducible method for the acquisition of tumor measurements [48]. The traditional WHO criteria are based on bidimensional thickness measurements that are more suitable for tumors that are relatively spherical, and the RECIST protocol calls for measurement of the longest lesion diameter, which is ill-defined for mesothelioma given its peripheral rind-like growth.

The modified RECIST protocol calls for linear tumor thickness measurements perpendicular to the inner surface of the thoracic cavity made at up to six locations ("measurement sites") in the chest; two thickness measurements may be made on each of three CT sections separated axially by at least 10 mm. Minimal measurable disease was set at 10 mm in the initial modified RECIST guidelines in accordance with the RECIST guidelines; the recently

updated version of the modified RECIST guidelines defined minimally measurable mesothelioma tumor thickness as 7 mm [47]. These (up to six) measurements are summed to provide an image-based estimate of tumor bulk at a given point in a patient's treatment. Patients are classified on the basis of the relative change in these summed thickness measurements between scans according to the RECIST classification criteria [49]. A patient thus will be categorized (1) as "partial response" (PR) if the summed thickness of disease has decreased by at least 30% from the baseline scan summed tumor thickness, (2) as "progressive disease" (PD) if the summed thickness measurement has increased by at least 20% from the nadir of the summed measurements of prior scans, and (3) as "stable disease" (SD) if the summed thickness has neither sufficiently decreased from baseline to be classified as PR nor sufficiently increased from the nadir of prior measurements to be classified as PD. Complete response indicates a complete disappearance of pleural and non-pleural lesions, which is rare in mesothelioma. Furthermore, the appearance of a new non-pleural lesion or an unequivocal new focus of pleural thickening that exceeds the minimum measurable size as defined above shall be considered PD under the recently updated version of the modified RECIST guidelines [47].

1.1.7 Tumor Volumetry for Mesothelioma Evaluation

Mesothelioma tumor volume has been found to correlate with patient survival for patients undergoing multi-modality therapy; furthermore, image-based tumor volume has been proposed as a marker for mesothelioma tumor response to treatment and as a part of an image-based staging system for this disease [50–55]. Three-dimensional imaging modalities, such as CT and MRI, allow researchers and clinicians to acquire image-based measurements of tumor volume. The following discussion on mesothelioma volumetric segmentation will focus on CT-based segmentation studies due to the scarcity of published studies on mesothelioma segmentation with MRI scans and the status of CT as the main imaging modality for clinical image-based evaluation of mesothelioma in the US. Methods to obtain image-based volumet-

ric measurements can be broadly classified into three categories: manual, semi-automated (i.e., partially computer-assisted), and automated methods.

Radiologist-obtained manual volumetric measurement is considered the "gold standard" for mesothelioma tumor and other lesions. This manner of volumetry is the most time-consuming and least practical for clinical applications when compared with semi-automated and automated methods; however, radiologists' expertise is required to validate and/or evaluate the performance of more time-efficient computerized methods for the task of disease segmentation for mesothelioma and other tumors. Evaluation of segmentation performance of computerized methods in this manner typically relies on expert-provided reference tumor segmentations on individual CT sections, selected at random from whole CT scans [56, 57].

A single study has assessed the correlation of radiologic and pathologic tumor volume in mesothelioma by evaluating the agreement between CT-based measurements of tumor volume acquired by a radiologist and pathologic measurements of tumor volume after a surgical tumor resection [58]. This study found a modest correlation (correlation coefficient $R^2 = 0.66$) between CT-based tumor volume and tumor specimen volume. The authors proposed that this relatively low correlation could be partly due to the challenges of visually separating tumor from adjacent structures in CT scans and partly due to the limitations of pathologic tumor volume measurement. The results of this study point to a knowledge gap in researchers' understanding of the ability of image-based tumor volume measurements to capture the pathologic volume of tumor in mesothelioma.

Labby et al. [59] investigated the interobserver variability of tumor area measurements in mesothelioma for the assessment of therapy response through the comparison of tumor segmentations provided by five radiologists on the CT scans of 31 patients. Three axial CT sections were pre-selected for each scan of the study in the same manner as when linear thickness measurements are made according to the modified RECIST protocol. This study found that the 95% confidence interval for relative interobserver variability of per-section area and summed area across the three sections both spanned a range of approximately 300%,

and the variability among observers was found to be a significant component of the total variability in area measurements. The authors observed that the interobserver variability did not decrease for measurements in which the area across the three sections of a scan were summed, which would be expected if the differences between observers were due to random fluctuations in the contours. This result suggests that the radiologists who participated in the study had different approaches to the task of tumor contouring or different perceptions of tumor boundaries [59].

Semi-automated methods for tumor volumetry in mesothelioma generally rely on observer-provided segmentations of tumor on selected sections as input; the algorithm subsequently applies an interpolation method to generate tumor segmentations on sections for which user input was not provided. Such methods have been used for the acquisition of image-based measurements of tumor volume for studies related to the assessment of patient survival, staging and response evaluation in mesothelioma [50–55]. Semi-automated methods for tumor volumetry are inherently less time-consuming than manual methods and as such provide a more feasible method over the fully manual approach for tumor volume acquisition in the research setting. The "quality" of semi-automatically acquired volumetric segmentations in mesothelioma may be called into question, however, as interpolation methods used in commercial software packages for this task are often unclear on the algorithmic implementation or overly simple (e.g., linear interpolation), and it is unclear how the considerable interobserver variability of mesothelioma tumor area measurements will affect volume measurements acquired by different observers using such a method [59].

The development of automated methods for mesothelioma tumor segmentation is a challenging task due to (1) the overlap in pixel values of tumor and adjacent soft tissue (including pleural fluid and collapsed lung) on CT, (2) the variability in tumor morphology between patients and between scans of the same patient, and (3) the variability in what experts perceive to be "tumor" on CT scans [18, 58, 59]. Sensakovic et al. [56] published a study on the automated segmentation of mesothelioma tumor on CT scans that employed a step-wise pro-

cedure to isolate the pleural space through segmentation of the lungs and the thoracic cavity before applying a machine learning-based classifier to classify the pixels within the pleural space as tumor and non-tumorous tissue. The method presented in this study achieved an average Dice similarity coefficient (DSC) value of approximately 0.65 when comparing computerized segmentations with radiologist-provided segmentations on an independent test set of 66 CT sections (DSC is a measure of overlap of two segmentations under comparison that takes values between 0, indicating no overlap, and 1, indicating complete overlap). The average Dice similarity coefficient when comparing observers' segmentations was found to be approximately 0.68; the difference between these values did not achieve statistical significance [56].

1.1.8 Issues Related to the Image-Based Evaluation of Mesothelioma

The morphology and growth patterns of mesothelioma tumor, combined with the low contrast of tumor and adjacent soft tissue and pleural effusion, complicate the acquisition of reliable estimates of tumor extent when compared with solid tumors that exhibit approximately spherical growth. Other considerations for the image-based evaluation of mesothelioma, apart from the inherent difficulty of separating disease from benign tissue, include the choice of measurement sites on a patient's baseline scan, the orientation of tumor thickness measurements relative to the lung and chest wall, and consistency in scan-to-scan locations of measurement sites. These aspects of image-based evaluation could have an impact on a patient's treatment management – as an example, clinical trials in mesothelioma and other cancers commonly rely on scan-to-scan image-based response classification to determine whether patients are eligible to stay on the treatment offered through the trial. In the case of mesothelioma, questions have been raised regarding the applicability of the response classification criteria of the RECIST guidelines to this disease [60].

The modified RECIST protocol introduced the concept of "measurement sites" for the image-based evaluation of change in tumor bulk between scans for mesothelioma patients and

aimed to standardize as many aspects of the measurement process for this complex disease as possible [46, 47]. Modified RECIST allows for up to six measurements of tumor thickness at baseline, oriented perpendicularly to the tangent of the outer lining of the pleura, on three CT sections separated by at least 1 cm craniocaudally, with up to two measurements per section [47]. This aspect of the measurement protocol for mesothelioma was intended to replace the measurement of longest lesion diameter as dictated by the RECIST guidelines; the longest diameter of a rind-like tumor characteristic of mesothelioma is both ill-defined (see Fig. 1.2) and does not directly correlate with tumor volume.

According to the modified RECIST guidelines for mesothelioma, each measurement site of tumor thickness must satisfy the criterion for minimally measurable disease, and measurements for subsequent follow-up scans should be made at the measurement sites chosen at baseline. To encourage reproducible measurements, the guidelines call for selection of measured tumor to be made "at the level of anatomic landmarks" (e.g., bony structures or the descending aorta); however, this consideration may not align with tumor locations that are likely to be of "clinical significance" due to the extent of tumor bulk or anatomic location. Furthermore, mesothelioma tumor appearance on axial CT sections may change due to patient respiration and positioning – to alleviate these factors, modified RECIST suggests that measurement sites be located between the aortic arch and the left atrium, if possible [47]. The development of a reliable method for automated tumor volumetry (see Section 1.1.7) could alleviate this reliance on the choice of measurement sites at baseline.

Other factors can affect the measurement of tumor thickness on a CT scan. In a study on the variability of mesothelioma tumor thickness measurements, the 95% limits of agreement for relative interobserver agreement in distinct measurements were found to span a range of 30% [61]. This study excluded the effect of measurement site selection; for patients who present with extensive tumor or multiple foci of pleural thickening at baseline, differences in the choice of tumor measurement sites would likely increase the variability in tumor thickness measurements. A study that investigated the minimal measurable thickness of mesothelioma

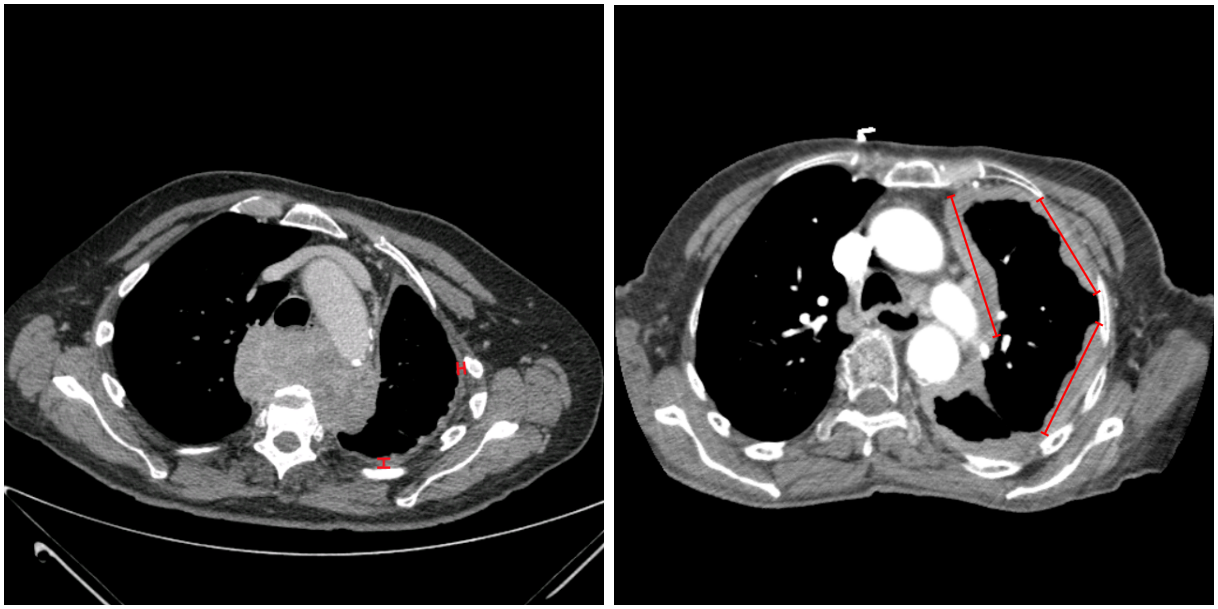

Figure 1.2: Axial CT section showing linear thickness measurements made according to the modified RECIST guidelines (left); three hypothetical measurements of what could be perceived as the "longest diameter" of a mesothelioma lesion on an axial CT section of a mesothelioma patient (right).

tumor on baseline CT scans indicated a consistent bias to longer or shorter measurements for some of the observers, suggesting that the same observer should perform all image-based response evaluations for a given clinical trial [62].

The validity of the use of summed linear thickness measurements to evaluate the tumor bulk of a non-spherical disease such as mesothelioma, and the applicability of the RECIST response classification criteria (which were initially derived for use with spherical tumors) to mesothelioma patients has been put into question. A study that investigated the use of aspherical geometric tumor models to evaluate the response classification criteria for mesothelioma demonstrated that the use of the RECIST criteria with modified RECIST might over-classify mesothelioma patients into the PR and PD categories [63]. Labby et al. [60] developed mesothelioma-specific response classification criteria for linear thickness measurements made according to modified RECIST through an optimization-based approach. The authors found that setting a relative change in summed thickness of -64% from baseline as the criterion for classification into the PR category and a 50% increase in summed tumor

thickness from the nadir measurement for PD classification improved the correlation between image-based response and patient survival in their cohort of 78 patients. The results of these two studies suggest that alternative methods for the evaluation of treatment response and the response categorization of mesothelioma patients should be investigated further.

As outlined in Section 1.1.5, novel therapeutic strategies are under active development for mesothelioma, which could call for new and more sensitive methods to measure tumor response to accurately evaluate these treatments. Some of these treatments for mesothelioma target VEGF, a protein that stimulates the formation of tumor blood vessels. Dynamic imaging modalities that allow for the tracking of tumor metabolism and physiologic change, in addition to morphologic change (i.e., change in tumor bulk), could prove to be useful for the evaluation of treatment response for these treatments [64]. Immunotherapeutic drugs aim to strengthen the response of the patient's own immune system to tumor cells; this type of treatment has been found to produce patterns of response that differ from traditional cytotoxic cancer therapies. In some cases, initial increase in the size of tumor lesions is perceived after the initiation of immunotherapy followed by a subsequent decrease in apparent tumor burden – this pattern of response to immunotherapeutics has been termed "pseudo-progression" and has prompted the development of immunotherapy-specific guidelines for the evaluation of tumor response to treatment [65, 66].

1.2 Dynamic Computed Tomography

Dynamic contrast-enhanced computed tomography (DCE-CT) is a functional imaging modality, which is a class of imaging modalities that provides information on tumor physiology; other functional modalities include PET/CT and DCE-MRI. DCE-CT tracks the uptake of iodinated contrast material in tissues through a sequence of CT acquisitions and was initially proposed by Axel [67] as a non-invasive method for the measurement of cerebral blood flow in the brain. The dynamic component of this method can be incorporated into conventional contrast-enhanced CT protocols. DCE-CT has been used to directly assess

tumor vasculature and angiogenesis in combination with the anatomic detail of CT and has been used to evaluate tumor response to cancer treatment and outcome prediction for solid tumors [68–72].

The underlying principle of DCE-CT is the tracking of tissue attenuation during and after the intravenous injection of iodinated contrast media. Tissue attenuation (or enhancement), as measured by the Hounsfield Unit (HU) value of a given volumetric pixel (voxel) on a reconstructed CT scan, increases linearly with iodine concentration [67]. The repeated CT imaging of an anatomic region immediately before and following the start of injection of high-attenuating contrast media allows for the tracking of increase in attenuation due to tissue uptake of contrast. Contrast distributes predominantly to the intravascular and extracellular spaces, and hemodynamic parameters may be estimated from the measurement of tissue attenuation through the application of kinetic two-compartmental models. In general, this process requires the measurement of an "arterial input function," which is the time-attenuation curve of an artery that supplies blood to the tissue of interest. Figure 1.3 shows the generic form of such time-attenuation curves for an arterial input function and tissue of interest – the contrast enhancement is evaluated over "baseline," which is taken to be the enhancement as measured on the first DCE-CT acquisition prior to injection of contrast [68, 73]. Using the arterial input function for the calculation of hemodynamic parameters in DCE-CT corrects for differences in cardiac function between patients, which may confound interpatient comparisons of tissue enhancement on regular contrast-enhanced CT [69].

Several kinetic models allow for the calculation of hemodynamic parameters using DCE-CT measurements. One assumption of these models is that there exists a single source of blood supply or, in the case of multiple supplies of blood, that contrast material concentration behaves identically within blood supplies across dynamic acquisitions; furthermore, it is assumed that contrast media mixes completely with blood [68, 69]. For the purposes of this work, the discussion will focus on kinetic models based on the Fick principle [69]; other

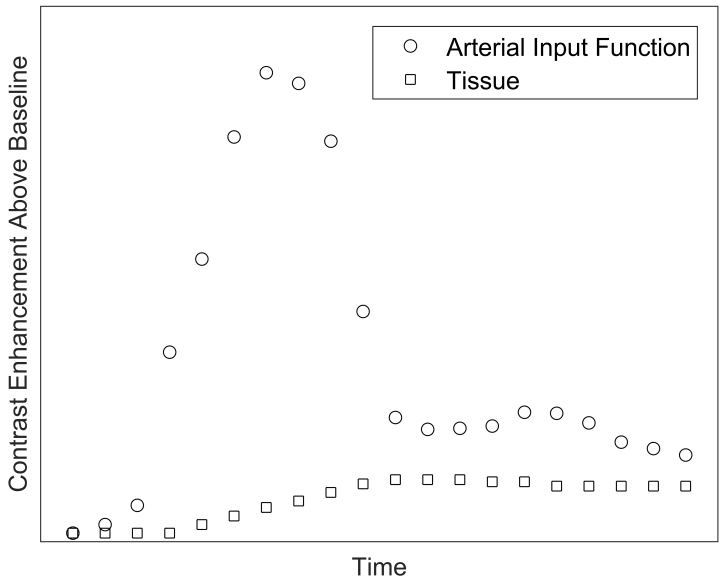

Figure 1.3: Time-attenuation curve of a DCE-CT protocol; contrast enhancement over baseline as a function of time for an arterial input function and tissue of interest.

models, such as those based on the Patlak principle, have been described elsewhere [69, 74, 75]. Hemodynamic parameters acquired with DCE-CT have been used as biomarkers in chemotherapy drug trials for a range of cancers [73]. Hemodynamic parameters that can be estimated through DCE-CT analysis include blood flow per volume of tissue (BF) and tissue blood volume (BV), which is the fraction of tissue that is taken up by blood. BF has been formulated according to the "slope method" as the ratio of the maximum slope of enhancement $Q(t)$ in tissue over baseline (i.e., enhancement prior to injection of contrast) and the maximum of the arterial enhancement $C_a(t)$ over baseline [76]:

$$\mathrm{BF} = \frac{\left[\frac{dQ(t)}{dt}\right]_{\max}}{C_a(t)|_{\max}} \quad . \tag{1.1}$$

BV can be approximated with DCE-CT as the ratio of peak tissue enhancement over baseline to the maximum arterial enhancement over baseline [77, 78]. BV has been found to correlate with VEGF expression and density of tumor microvessels for a variety of tumors [73].

Technical considerations for DCE-CT include the volume, concentration and rate of injection of contrast media used with the acquisition. DCE-CT analysis will benefit from rapid injection of lower volumes of highly concentrated contrast material; however, the range of feasible values of these parameters will be reduced for dynamic imaging incorporated into standard clinical CT protocols [78]. Sources of error in DCE-CT studies include image noise, partial volume effects and image misregistration due to patient movement between dynamic acquisitions [69].

1.3 Texture Analysis

Texture analysis is a method for the quantitative analysis of digital images that can be applied to the image-based classification of tumors and other patient anatomy. In general, this method involves the extraction of a collection of numerical values based on mathematically defined "texture features" that describe the patterns of appearance and correlation of pixel values in given regions of interest (ROIs) within an image. These feature values are used as input to a classifier to assist in decisions based on image information that might not be apparent to human observers. The texture analysis of biomedical images from a variety of imaging modalities has long been of interest as a component in computer-aided diagnosis (CAD) systems, which are computerized systems that aim to assist physicians in answering clinical questions, such as the estimation of tumor likelihood of malignancy [79–84]. The extraction of a high number of texture features for the classification and characterization of radiologic images has been termed "radiomics" [85, 86].

Texture features can be used to characterize a region within an image and, in the context of medical imaging, have been used to characterize and classify ROIs that may contain a lesion or other pathology [87, 88]. Spatial heterogeneity within cancerous tumors has been found to indicate poor patient prognosis, and cancer treatment may affect such intratumor heterogeneity [89, 90]. CT textural analysis has been applied to the assessment of tumor response to various types of treatment in solid tumors including hepatocellular carcinoma,

non-small cell lung cancer and renal cell carcinoma [91–94].

1.3.1 Categories of Texture Features

Features computed through texture analysis can be divided into several broad categories. For the purposes of this work, the following discussion will focus on two categories of texture features, 1) histogram-based features and 2) features based on the gray level co-occurrence matrix (GLCM), along with shape-based features [95–97]. Other feature categories include fractal features, gray level run-length coding and features based on Fourier transform analysis [95, 96, 98, 99].

Histogram-based feature values are calculated directly from the histogram of pixel values of an ROI. This class of features does not take the spatial relation of pixel values into account and includes statistical features derived from the histogram of an ROI such as the standard deviation, mean, median, and 95th percentile of pixel values, as well as features that measure the "randomness" and spread of the ROI's histogram, such as kurtosis and energy [96].

Features that take into account spatial relation of grayscale pixel values include features calculated from the statistics of the GLCM, first proposed for texture-based image classification by Haralick et al. [95]. The GLCM is an $N \times N$ matrix, where N represents the number of gray levels in an image region; the gray levels of the region may be binned before the calculation of the GLCM or limited to a range of gray levels out of the entire range of possible gray level values (e.g., the 256 possible pixel values for an image stored in 8-bit integer format). Each GLCM is defined by an angle θ and pixel distance d; the value in the i-th row and j-th column of the GLCM represents how many times the pixel value represented by the i-th row appears in the ROI at an angle θ and distance d from the pixel value represented by the j-th column [100]. After the GLCM for a given θ and d has been constructed, texture features that relate to homogeneity, contrast and structure within the image region are extracted from the GLCM through a set of equations [95]. To obtain directionally independent feature values, an aggregate value for an ROI may be calculated by taking the average of

feature values calculated from GLCMs associated with several values of θ [101].

1.3.2 Texture Feature Classification

The machine learning-based classification of image regions from texture feature values may be accomplished through a variety of methods. "Supervised" classification schemes involve the development of a computerized model for the classification of input data (in the case of texture analysis, a feature vector) and a known reference classification ("class label") of the input samples [102, 103]. "Unsupervised" methods aim to identify trends in the input data without the use of the reference class labels and are in some cases inappropriate for the unbiased and generalizable classification of high-dimensional data [104]; for the purposes of this work, the following discussion will focus on supervised methods.

Early texture feature-based CAD systems employed simple rule-based schemes for lesion classification; these schemes have, for example, shown success in classifying benign and malignant masses in mammograms [83, 105]. For other classification tasks in this field, more advanced classification methods, such as linear discriminant analysis (LDA), artificial neural networks (ANNs), and support-vector machines (SVMs) have been used for texture feature-based classification of pathologies [88, 96, 106].

Texture features obtained from ROIs in medical images can number in the tens or hundreds, even outnumbering the number of samples used for training and validating the model [107]. Considerations when training a machine learning model to classify such high-dimensional data include (1) preventing the classifier from overfitting to the training set, which limits the generalizability of the model and (2) ensuring the trained classifier is tested on an independent test set, in order to obtain an unbiased estimate of the model's classification performance [104, 105].

1.4 Deep Convolutional Neural Networks

Machine learning-based classifiers are computerized models that are "trained" to classify input data into groups or classes (e.g., benign and malignant lesions in the context of medical imaging). The training of such models is usually achieved through the minimization of a loss function, which provides a measure of the discrepancy between the classification output of the model and a given reference classification. The principal aim of machine learning-based classification models is to achieve a high classification performance on unseen datasets (i.e., high "generalizability"). Deep convolutional neural networks (CNNs) are a type of machine learning classifier that are commonly applied to the classification of images. For the purposes of this work, the following discussion will focus on "feedforward" deep CNNs, i.e., networks for which the input data flows only in one direction.

1.4.1 Building Blocks of Deep Convolutional Neural Networks

Artificial neural networks (ANNs) consist of a layer, or layers, of so-called "neurons" that take as input a vector $\vec{X}$ (which can be the original input to the network or output of the network's previous layer) and transform it through multiplication by a weight vector $\vec{W}$, a bias term b and a non-linear activation function σ, producing an output activation h_k:

$$h_k = \sigma\left(\vec{W}\vec{X} + b\right) \quad . \tag{1.2}$$

ANN are typically trained, through an optimization-based process, to classify samples of input data x into classes y. During this process, the weights and biases of the network are modified to reduce the discrepancy between the network's predicted classification output $\hat{y}$ and the reference output y; in general, the network classification $\hat{y}$ is produced by thresholding the activations h_k of the final layer of the network [108]. For the purposes of this work, the following discussion will focus on classification tasks for which there are only two possible output classes (e.g., "benign" and "malignant").

Deep CNNs are a subset of deep ANNs – the adjective "deep" indicating that these networks have more than one so-called "hidden" layer of neurons between the input and output layers of the network. The "depth" of a layer indicates how many distinct layers of computation (e.g., convolution or downsampling) the data to that layer have undergone. The general classification strategy of deep CNNs can be described by Eq. (1.2); however, CNNs and ANNs differ in their propagation of input data through the network. Deep CNNs treat the weight $\vec{W}$ in Eq. (1.2) as an n-dimensional weight matrix $\mathbf{W}$ that forms a convolutional "filter" of a given dimension $\vec{d}$. This approach to the extraction of information from the input data makes deep CNNs highly efficient for the classification of images, which are generally stored and viewed as two-dimensional matrices [108]. The following discussion will focus on the two-dimensional case, for which $n = 2$.

Each layer of a deep CNN produces a so-called "feature map" through successive element-wise multiplication and summation (i.e., convolution) of the filter $\mathbf{W}$ and the input data to the layer; the filter is "scanned" across the input and produces an activation at each distinct position of the layer input. Figure 1.4 shows a convolution operation for a 3×3 filter of a deep CNN. In the case shown in Fig. 1.4, only "valid" convolutions, i.e., positions at which the convolution filter is completely contained within the layer input, are stored in the resulting feature map; however, the input to the layer may be padded with zeros or the nearest value of the input to generate feature maps of the same size as the layer's input. As in Eq. (1.2), a bias term b is added and a non-linear function σ is applied to the resulting value to produce the final feature map of activations. Non-linear activation functions include the rectified linear function (ReLU):

$$\sigma(x) = \max(0, x) \tag{1.3}$$

and sigmoid-type functions, an example of which is the logistic function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad . \tag{1.4}$$

The functions defined in Eqs. (1.3) and (1.4) are shown in Fig. 1.5.

Other types of layers in deep CNNs include downsampling layers, which allow the network to incorporate larger-scale features of the input without increasing the dimension $\vec{d}$ of the filters used to generate the feature maps, and fully-connected layers, which are layers of neurons, analogous to ANNs, that are often used as the final set of layers of deep CNNs for the prediction of global classifications of images or image regions [108]. A dropout layer will set individual activation values to zero in its input feature maps during training at a given probability, which has been shown to reduce "overfitting" to the training set (i.e., loss of generalizability) in deep neural networks [109]. The collection of layers and layer connections that define a deep CNN is known as the "model architecture" of the network.

During the training process of a feedforward deep CNN, input samples from the training set will sequentially undergo the mathematical operations at different depths of the network, producing as output the predicted classification $\hat{y}$ for the input. The predicted classification of the network is subsequently compared with the reference classification y by calculating the so-called loss function of the model. A common loss function employed in deep learning is the cross-entropy, which is based on the idea of maximum likelihood estimation and can be viewed as a measure of dissimilarity between the distribution of predicted classifications and the distribution of reference classifications on the training set [108]. For a two-class problem, the binary cross-entropy loss function can be defined as:

$$L(t_i, p_i) = -\left(t_i \log\left(p_i\right) + \left(1 - t_i\right) \log\left(1 - p_i\right)\right) \quad , \tag{1.5}$$

where t_i takes the value 0 if sample i is of the "true" class (e.g., malignant) and the value 1 if the sample i is of the "false" class (e.g., benign), and p_i is the predicted probability of input sample i belonging to the "true" class. The back-propagation algorithm [110] allows the calculation of the gradient of L with respect to the weights and biases of the network. The stochastic gradient descent algorithm is most commonly used to update the weights and

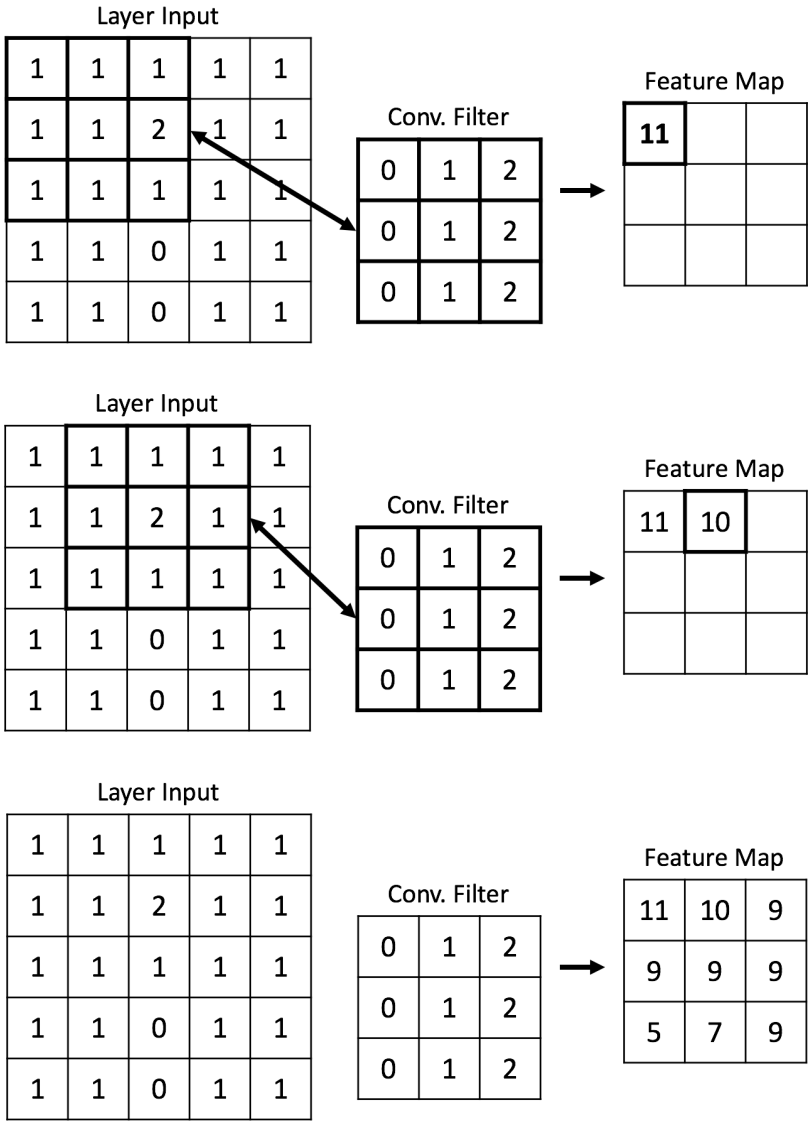

Figure 1.4: Feature map calculation at a layer of a deep CNN showing (top) the element-wise multiplication and summation that produces the top left value of the feature map, (middle) the element-wise multiplication and summation that produces the top middle value of the feature map, and (bottom) the final feature map result. Here, only "valid" convolutions are allowed, for which the convolution filter is contained completely within the layer input, which results in a feature map that is reduced in width and height by one pixel when compared with the original size of the layer input. For clarity, the bias b and non-linear function σ that produce the final feature map of layer activations are not shown.

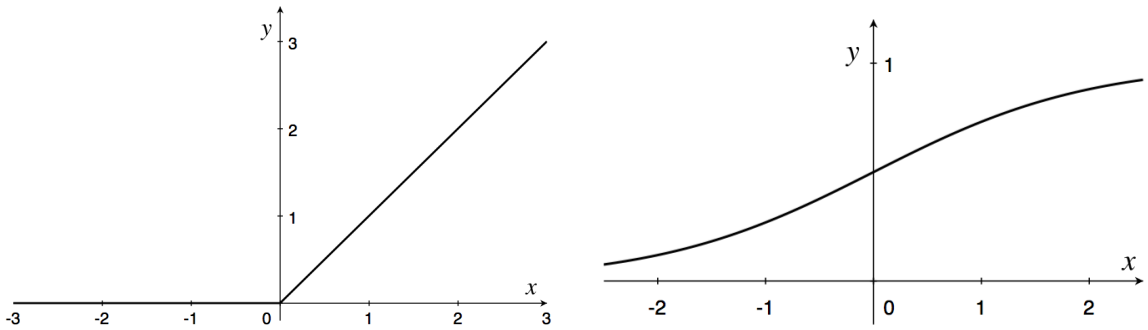

Figure 1.5: Non-linear activation functions commonly used in deep CNNs, as defined in Eqs. (1.3) and (1.4): (left) a ReLU activation function; (right) the logistic function, a sigmoid-type activation function.

biases of the deep CNN using the information learned through back-propagation of the loss. The parameters of the network can be updated according to the loss calculated after a single input sample has been fed through the network or using larger "batches" of samples, which reduces the noise in the updates to the network parameters [108, 111].

A layer at a given depth of a deep CNN will, in general, have multiple convolutional filters that are applied to the input to the layer; a layer of 64 filters will thus produce 64 feature maps as its output. The filter dimensions $\vec{d}$ at each depth of the network, the number of filters at each depth of the network, and the dropout probability are some of the many hyperparameters (i.e., design parameters) that affect the performance of deep CNNs for tasks in medical imaging and other areas of study. The optimization of these hyperparameters is generally achieved by evaluating the performance of the trained network on a validation set that has not been used to guide the loss minimization of the network during training. This validation set should be different from the final test set on which the optimal deep CNN, following hyperparameter optimization, will be applied. This independent training, validation and testing methodology reduces the bias of the final predicted classification on the test set.

1.4.2 Convolutional Neural Networks in Medical Imaging

The 1990s saw the earliest investigations of CNNs for the purpose of medical image classification [112, 113]. In recent years, the availability of larger data sets, and the feasibility of larger model architectures due to increased computational power and better software infrastructure, has resulted in an increased interest in deep CNNs for tasks in image classification and segmentation, including in the field of medical imaging [114, 115]. The U-Net deep CNN architecture introduced by Ronneberger et al. [116] is a common deep CNN architecture used for the task of medical image segmentation [117]. Transfer learning, a technique in which the convolutional layers of a deep CNN model are initialized using weights and biases from a network that was previously trained on a larger natural image data set, has proved useful both to extract features and to enhance the training of networks for medical imaging-related classification tasks for which limited training data are available [88, 118, 119].

1.5 Outline

This dissertation presents a series of investigations that aim to augment the quantitative imaging-based methods available to researchers for investigations related to mesothelioma and to clinicians for the management of mesothelioma patients.

Chapter 2 presents the results of a prospective pilot study that investigated the use of DCE-CT for the assessment of treatment response through comparisons of hemodynamic parameter changes between scans of patients that were on treatment and patients that were off treatment. DCE-CT can be readily incorporated with standard clinically indicated imaging at a relatively low additional radiation dose to the patient and has been used in clinical trials to demonstrate drug effects in cancers such as non-small cell lung cancer and colorectal cancer. The hypothesis of this study was that differences in hemodynamic change between the two patient groups as measured by DCE-CT would support the utility of this imaging method for detecting changes in tumor function that are not necessarily evident on

standard imaging through changes in tumor morphology.

Chapter 3 implements a deep CNN-based method for the segmentation of mesothelioma on CT scans. Image-based tumor volume has been shown to correlate with patient survival in mesothelioma and has been suggested by researchers for use in the image-based staging and treatment management of this disease. The development of automated computerized techniques for the volumetric segmentation of mesothelioma presents a time-efficient method for the acquisition of tumor volume in both research and clinical settings. This study will compare a deep CNN-based method with a previously published step-wise segmentation method for mesothelioma and hypothesizes that, given a large training set of tumor contours, the advanced combination of large-scale and smaller-scale features captured in deep neural networks will allow a deep CNN to achieve state-of-the-art performance for the task of mesothelioma tumor segmentation on CT scans.

Chapter 4 presents a deep CNN-based method for the improved differentiation of mesothelioma tumor and pleural effusion on CT scans. Pleural fluid often accompanies mesothelioma and has considerable overlap in pixel values with mesothelioma tumor on CT, and erroneous inclusion of fluid within tumor contours will confound the estimation of tumor bulk acquired through a volumetric segmentation method. The training set of CT scans that was used for the study presented in Chapter 3 will be augmented with CT sections that exhibited both mesothelioma tumor and pleural effusions, and a comprehensive strategy will be developed to evaluate the performance of this method during training. The segmentation performance of the optimal trained networks will be evaluated on a test set of radiologist-provided reference tumor segmentations on CT sections that exhibit both tumor and pleural fluid. This study hypothesizes that through a dedicated training set of CT sections that exhibit pleural effusions and a comprehensive evaluation strategy during training to select the optimal networks, the segmentation performance of this method will surpass that of the method presented in Chapter 3 on CT sections of mesothelioma patients that present with both tumor and pleural fluid.

Chapter 5 of this work presents a retrospective investigation of the potential utility of CT scans for the noninvasive diagnosis of mesothelioma tumor subtype. Tumor histologic subtype is the most significant prognostic factor in mesothelioma, with the epithelioid subtype carrying a better prognosis when compared with biphasic (i.e., mixed epithelioid and sarcomatoid) and sarcomatoid tumor subtypes. This factor can determine the avenues of treatment available to mesothelioma patients and whether a patient should be considered eligible for surgery. A single study has shown the potential of dynamic MR imaging in the differentiation of epithelioid and non-epithelioid subtypes. For the study presented in this chapter, a database of CT scans of patients that have undergone tumor resection at our institution was collected and texture features were extracted from contoured tumor for the image-based classification of epithelioid and non-epithelioid mesothelioma. This study hypothesizes that epithelioid and non-epithelioid mesothelioma present with subtle differences in appearance and texture on CT scans and that these differences might lend themselves to the noninvasive differentiation of these subtypes.

Chapter 6 will summarize the main conclusions that may be drawn from this work and make suggestions as to further avenues of inquiry that could build on the studies presented in this dissertation. This dissertation will hopefully spark further research on methods for the quantitative imaging of mesothelioma and serve to motivate further investigations into the role of these methods in the clinical management of mesothelioma patients.

CHAPTER 2

DYNAMIC CT FOR MESOTHELIOMA TREATMENT RESPONSE

2.1 Introduction

As discussed in Section 1.1.6, the standard clinical method to assess mesothelioma tumor response with CT scans is based on linear thickness measurements of tumor, made at up to six locations in the chest. Patients are subsequently classified into response categories according to the relative change in summed tumor thickness between treatment time points [46, 47]. Mesothelioma tumor is typically aspherical, however, which has cast doubt on the suitability of linear tumor thickness measurements as a surrogate for tumor bulk and on the applicability of the RECIST criteria to the classification of mesothelioma tumor response [63]. Novel molecularly targeted therapies are under active development for use in the treatment of mesothelioma, including antiangiogenic therapies [36, 120, 121]. Both immunotherapeutic and antioangiogenic agents used in the treatment of cancer have been found to produce radiologic tumor response incompatible with the RECIST criteria, giving rise to other proposed response criteria for these types of therapy [65, 122, 123].

In recent years, imaging techniques that provide direct information on tumor function have been researched widely as ways to provide earlier indicators of treatment efficacy. Chemotherapeutic treatments are known to alter the hemodynamic characteristics of a variety of cancers, and earlier indicators of tumor therapeutic response would prove valuable in the treatment of mesothelioma. As described in Section 1.2, DCE-CT is an imaging technique that involves the repeated acquisition of images during and following the intravenous injection of contrast media [68, 78]. DCE-CT adds a temporal component to a CT scan that allows for the tracking of blood flow to different tissues over the duration of an imaging procedure, through which information can be obtained on the hemodynamic properties of the tissues of interest. Studies have shown that hemodynamic parameters derived using

DCE-CT may be sensitive to the decrease in tumor vascularisation induced by both cytotoxic and targeted cancer therapies in various cancers [70]. DCE-CT has been proposed as an alternative to standard imaging-based methods of response assessment based on changes in tumor morphology and has recently been applied to the assessment of tumor response in a variety of cancers, including non-small cell lung cancer and hepatocellular carcinoma [71, 72]. The aim of this prospective imaging study[1] was to investigate the utility of hemodynamic information derived from DCE-CT parameter maps in the assessment of tumor response to treatment in mesothelioma patients.

2.2 Patient Cohort

Patients were eligible for this prospective study if they had been diagnosed with mesothelioma with pathologic confirmation and if they were scheduled to receive a clinically indicated standard contrast-enhanced chest CT scan. Patients were also required to have had tumor with a thickness of at least 1 cm located axially between the superior-most point of the diaphragm and the superior-most point of the aortic arch on a previous CT examination. This requirement ensured a sufficient volume of disease to allow for quantitative analysis. Furthermore, patients were only included if they were not considered candidates for surgery and had not undergone talc pleurodesis in the three months prior to the first DCE-CT scan; the inflammation and increased FDG-avidity of pleural tissue seen in patients following pleurodesis could have a confounding effect on DCE-CT-derived hemodynamic parameters [42, 44].

The final patient cohort included 14 patients, of whom nine were on chemotherapeutic treatment and five were on observation and not receiving any biologically active therapy (see Table 2.1). Patient accrual occurred from mid-2010 to late-2013, and the study was approved by the institutional review board (IRB) at the University of Chicago. All patients provided written consent for their participation. As of June 1, 2016, 13 out of 14 patients had died (median survival from diagnosis: 1225 days, range 281 - 2717 days).

1. This study has been published in a peer-reviewed journal [124].

Table 2.1: Patient Characteristics ($n = 14$).

Characteristic	N (%)
Sex	
Male	11 (79%)
Female	3 (21%)
Age at diagnosis (years)	
Median	76
Range	52-84
Histology	
Biphasic	2 (14%)
Desmoplastic	1 (7%)
Epithelioid	10 (71%)
Sarcomatoid	1 (7%)
Treatment combination	
No biologically active treatment	5 (36%)
ARQ 197	1 (7%)
Carboplatin, Pemetrexed	1 (7%)
CBP501, Cisplatin, Pemetrexed	1 (7%)
Cisplatin, Pemetrexed	2 (14%)
GDC-0980	1 (7%)
Pemetrexed	1 (7%)
Vorinostat	2 (14%)

2.3 DCE-CT Protocol

A Philips Brilliance iCT 256-slice scanner was used for all scans in this study, and DCE-CT acquisitions were integrated with the standard imaging protocol for clinically indicated chest CT scans for mesothelioma patients. The image-acquisition protocol included two dynamic contrast-enhanced phases, one prior to and one following a standard CT scan of the full chest. First, a non-enhanced scout scan was performed to localize the anatomic region to be dynamically imaged. The dynamic section of each patient was constrained to the anatomic region between the top of the aortic arch and the top of the diaphragm to reduce partial volume effects due to the oblique pleural surface near the lung apices and to ensure the dynamic imaging of an artery for DCE-CT parameter calculation. Furthermore, the dynamic section was chosen such that it included clearly visible tumor located as superior as possible to minimize the effects of respiratory motion. Iohexol contrast media (brand name Omnipaque, 350 mg/ml iodine concentration; Amersham Health, Princeton, NJ) was used in the standard volume (90 ml) for a full-chest CT scan followed by 30 ml of saline chaser. The injection rate was 6 ml/second and was increased relative to other DCE-CT studies to partially compensate for the increased volume of contrast used for the clinically indicated contrast-enhanced CT scans of the present study [125, 126].

A graphical representation of the imaging protocol of this study is shown in Fig. 2.1. The dynamic acquisition protocol consisted of an early phase of 20 scans with a 3-second temporal sampling and a later phase of 5 scans with a 5-second temporal sampling following the full chest CT scan. Image acquisition of the first dynamic phase began concurrently with the injection of contrast; the dynamically imaged section of each patient was acquired without moving the patient table, which limited its axial extent to the effective width of the CT detector (approximately 55 mm). Patients were instructed to use shallow breathing during the two dynamic phases to minimize motion artifacts and differences in patient position that could affect the pixel-wise tracking of contrast enhancement across the dynamic CT acquisitions. The start time of the second dynamic phase varied slightly across patients;

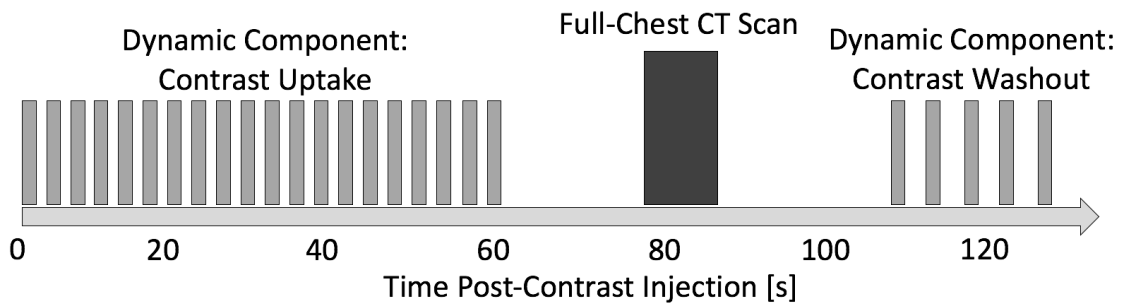

Figure 2.1: Graphical representation of the DCE-CT protocol of this study.

the median start time of the second dynamic phase was 114 seconds after injection (range: 108–213 seconds). The imaging parameters of the dynamic phases were set to 120 kVp with 100 mAs (0.33 s rotation time) for the initial 10 patients accrued for this study. The kVp setting was changed to 100 kVp for both scans of the last four patients included in the study, following an optimization study of exposure parameters with respect to image quality and patient dose. Reconstructions for the DCE-CT acquisitions were set at 3 mm slices using the standard kernel without edge enhancement.

Phantom studies indicated that the dynamic component of the imaging protocol would contribute approximately 1.5 times the dose of the standard chest CT scan, a finding based on the approximate linear scaling relationship between patient effective dose with CT and the reported dose-length product [127, 128]. Each patient underwent two DCE-CT scans at 60-day intervals on average (median: 50 days, range: 41–126 days) in conjunction with standard clinically indicated scans.

2.4 Image Postprocessing

DCE-CT data were analyzed using in-house software developed in Matlab (MathWorks, Natick, MA). Data acquired during the second dynamic phase were not used in the analysis due to the start-time variability across patients. Dynamic images were corrected for motion across dynamic time points by registering all axial sections at each individual acquisition of the dynamic phase to the 20th dynamic acquisition using the Advanced Neuroimaging

Tools (ANTs) software package [129]. Volumetric regions were manually constructed by an imaging scientist (E.G.) around all visible mesothelioma tumor on the dynamic sections of each patient. Temporal minimum- and maximum-intensity projection image sets across all dynamic acquisitions of each DCE-CT scan were used to exclude visible calcifications, blood vessels, collapsed lung, and pleural effusion from tumor regions. Fissural tumor, tumor within the hilar region, and tumor invading the chest wall were excluded by following expected pleural margins, using the ribs as reference. A volumetric region was constructed within the ascending aorta close to the level of bifurcation of the pulmonary artery to measure the arterial input function (AIF) required for the calculation of DCE-CT parameters. All regions were reviewed and validated by an experienced thoracic radiologist (C.S.). Figure 2.2 shows the time-attenuation curve for the aorta and a tumor ROI for a patient in this study, and Fig. 2.3 shows histograms of tumor enhancement at different dynamic acquisitions for both DCE-CT scans of a patient on treatment in this study.

2.5 Quantitative Analysis

2.5.1 DCE-CT Parameters

DCE-CT parameter maps of tissue blood flow (BF) and tissue blood volume (BV) were obtained using the maximum slope method. BF was calculated as the ratio of the maximum initial slope in the tissue and the maximum arterial enhancement over baseline [68]. For this study, we considered changes in BF between scans only for voxels of tissue that were actively perfusing; therefore, any voxels that showed no net enhancement across the dynamic acquisitions following the arrival of contrast in the ascending aorta were excluded from the calculation of BF. BV was calculated as the ratio of the maximum tissue enhancement over baseline and the maximum arterial enhancement over baseline, regardless of net enhancement across the dynamic acquisitions [77]. Figure 2.4 shows example BF and BV parameter maps from a patient from the treatment group of this study. Tumor DCE-CT parameter values

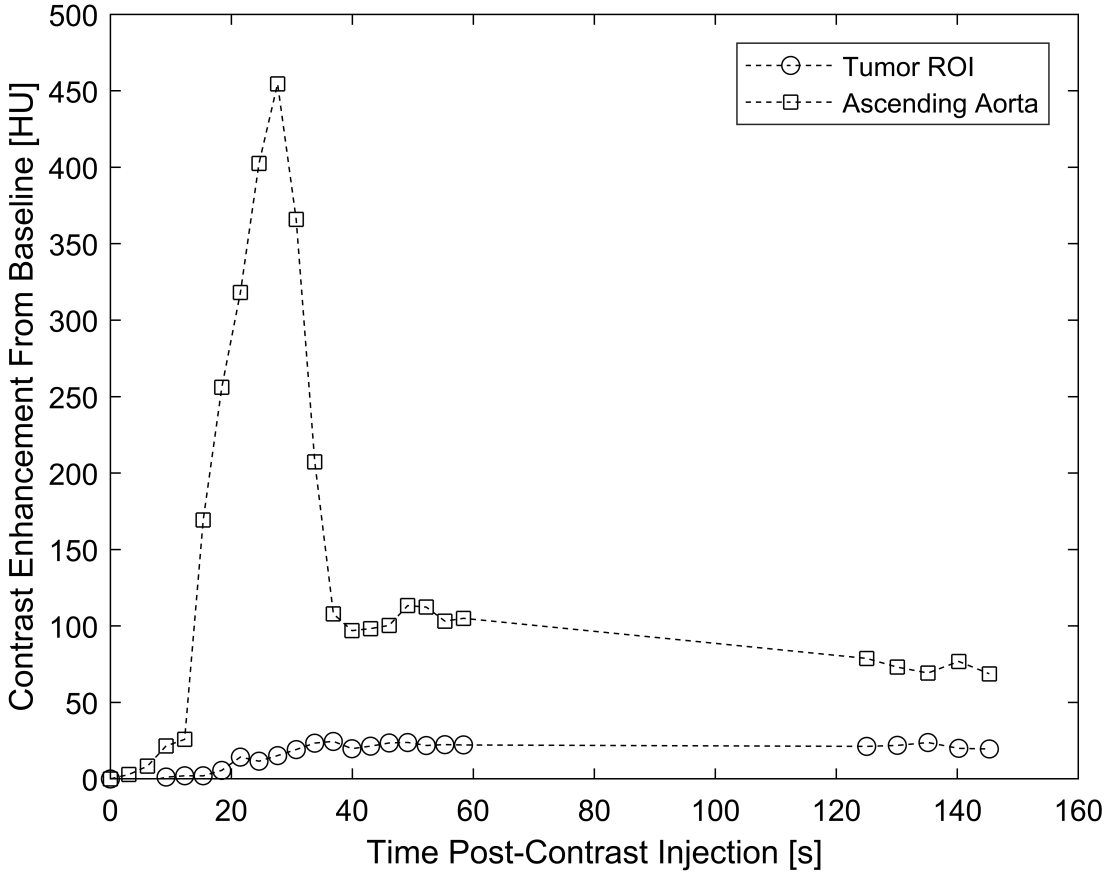

Figure 2.2: Contrast enhancement as a function of time post-contrast injection of the ascending aorta and a tumor ROI for a patient of this study.

were calculated as the mean value of the respective parameter map across all pixels within the volumetric tumor regions. Figure. 2.5 shows histograms of BF and BV values within tumor regions for both scans of a patient on treatment in this study.

To investigate the relationship between change in DCE-CT parameters and change in tumor bulk, summed linear thickness measurements were acquired by a radiologist (F.L.) using the full chest CT scan according to the modified RECIST guidelines [46].

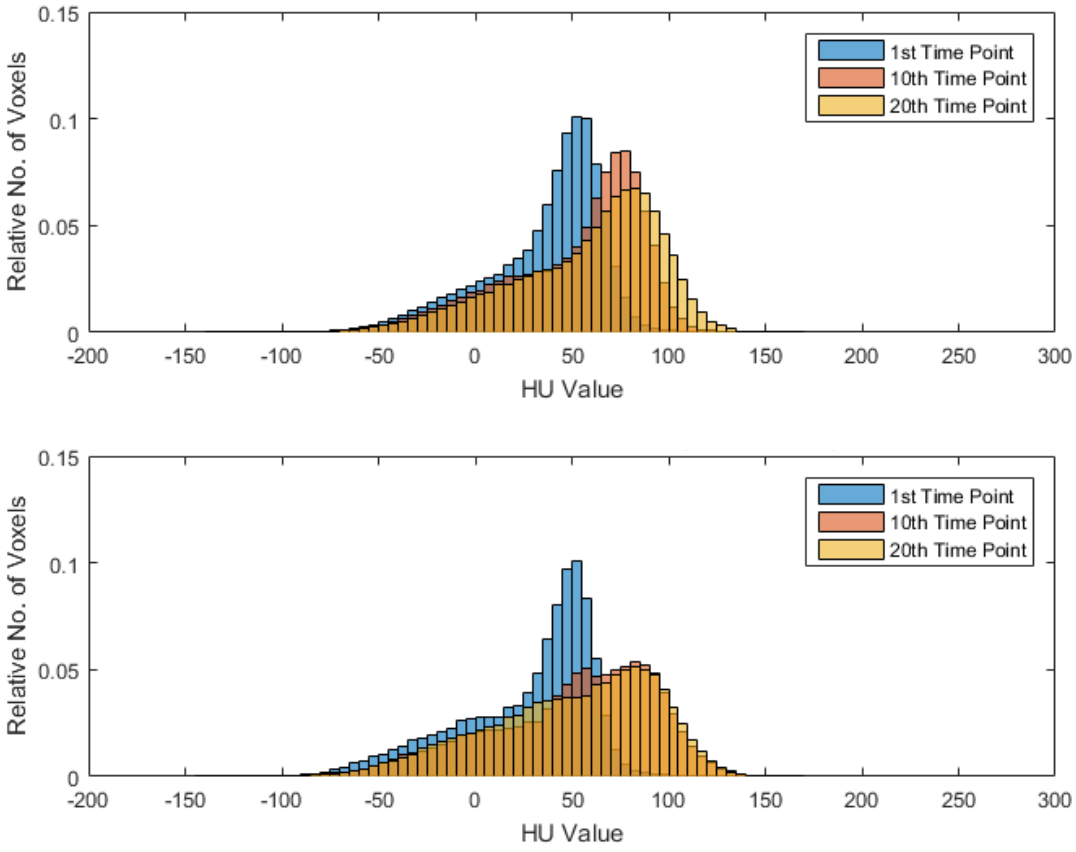

Figure 2.3: Histograms showing tumor enhancement at the 1st, 10th and 20th dynamic acquisitions for a patient on Vorinostat treatment in this study: (top) scan 1, (bottom) scan 2. Histograms at each dynamic acquisition are normalized to an area of 1.

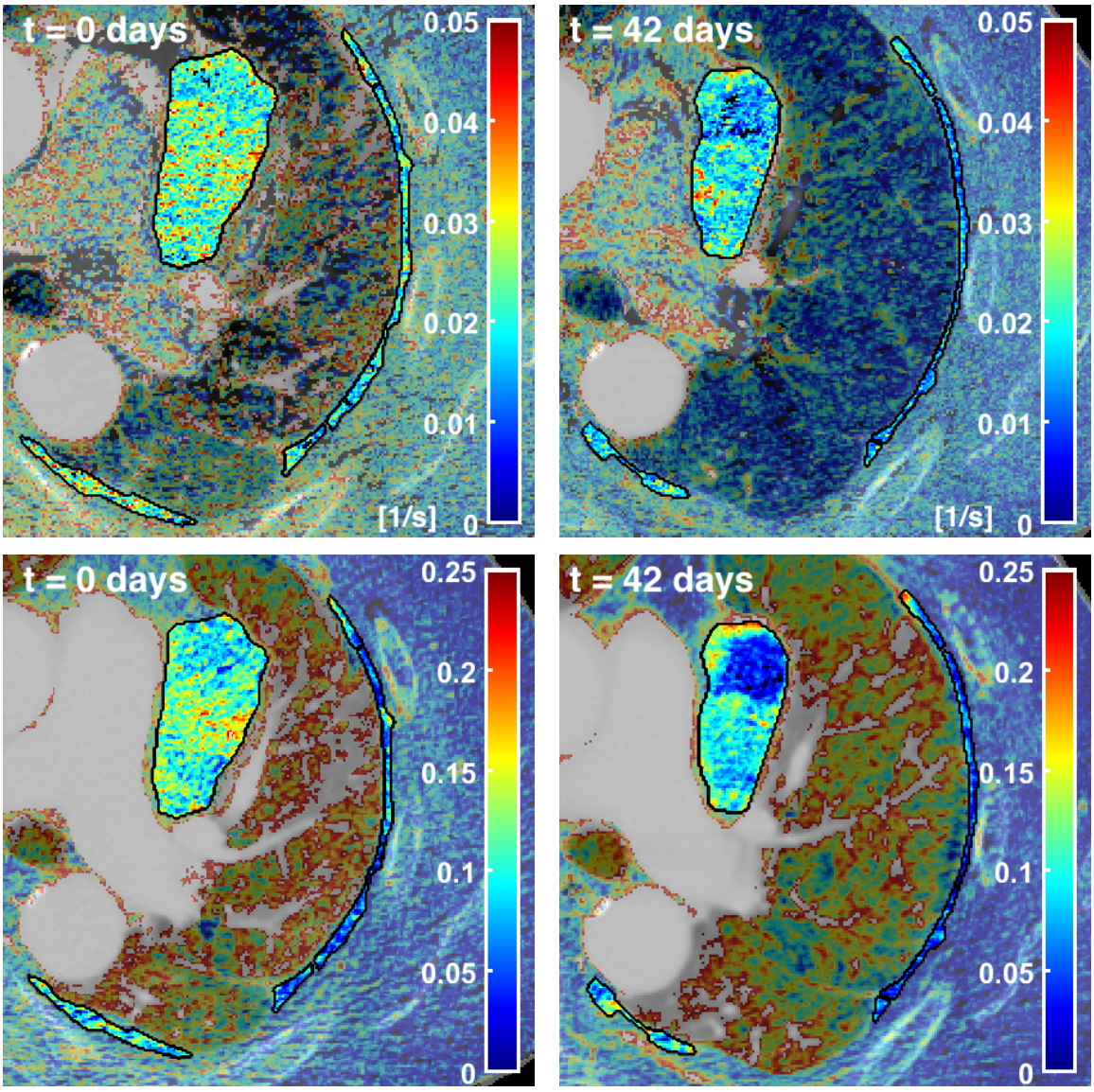

Figure 2.4: Example DCE-CT parameter maps of a patient on Vorinostat treatment in this study: (top, left) BF parameter map, scan 1; (top, right) BF parameter map, scan 2; (bottom, left) BV parameter map, scan 1; (bottom, right) BV parameter map, scan 2. Mesothelioma tumor regions shown with black solid lines. The parameter map outside tumor regions is partially transparent and overlaid on the corresponding grayscale CT section. This patient showed a 22% decrease in BF, an 8% decrease in BV, and a 10% decrease in modified RECIST tumor thickness measurements between scans.

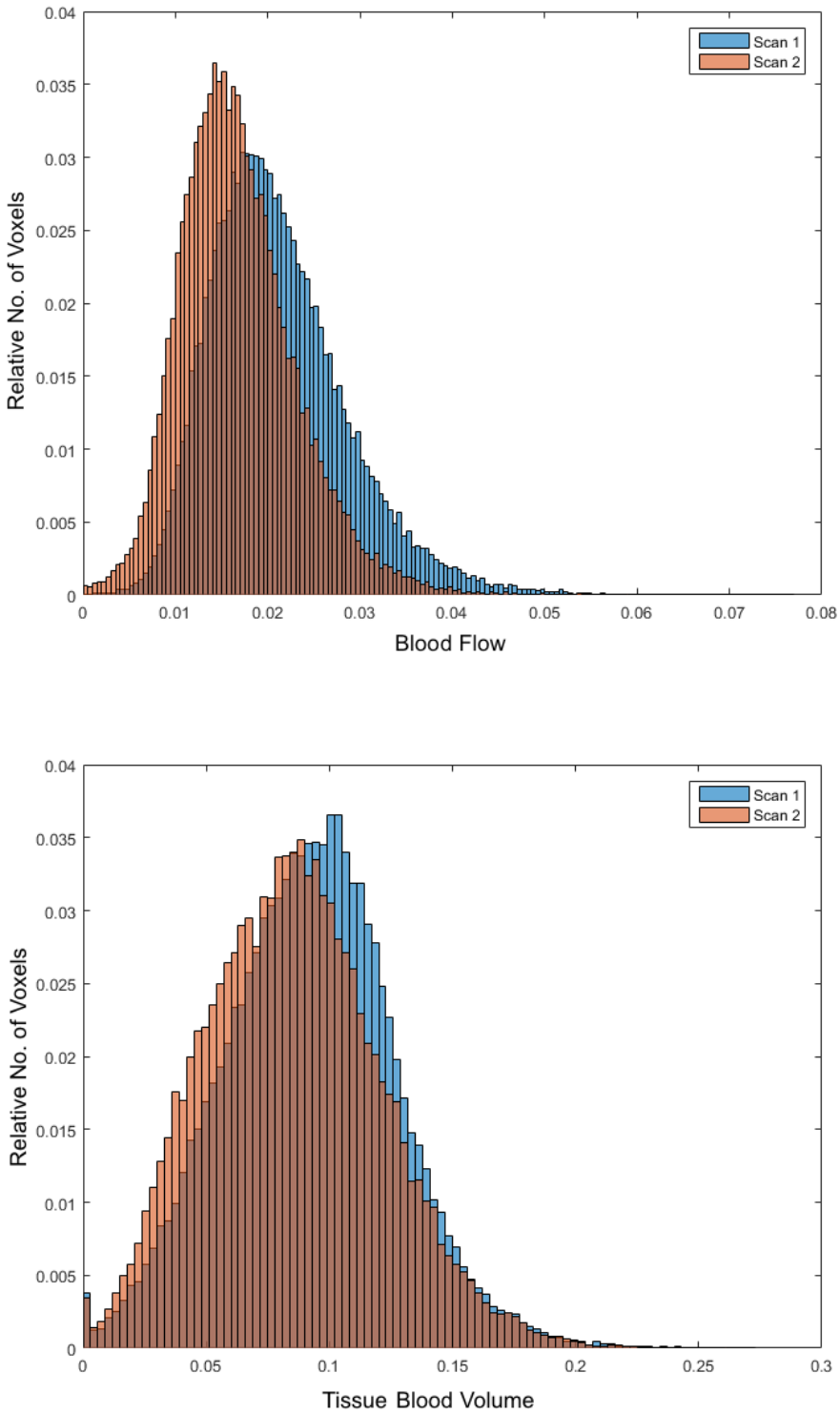

Figure 2.5: Histograms of DCE-CT parameter tumor values from scan 1 and scan 2 for a patient on Vorinostat treatment in this study: (top) BF, (bottom) BV. Histograms at each scan are normalized to an area of 1.

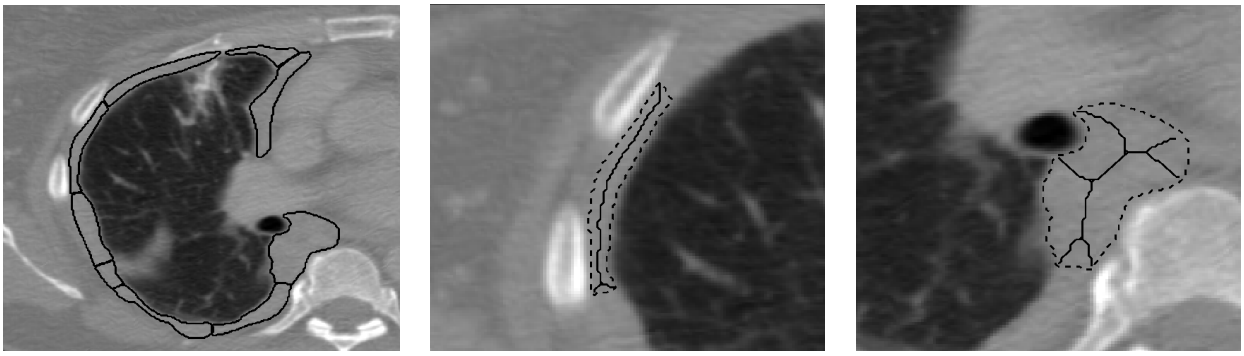

Figure 2.6: Mesothelioma tumor contoured on a single dynamic section of a DCE-CT scan: (left) all tumor regions, (middle) a lateral tumor region shown with original (dashed line) and skeletonized (solid line) tumor regions, (right) a medial tumor region shown with original (dashed line) and skeletonized (solid line) tumor regions.

2.5.2 Tumor Contour Variations

DCE-CT parameter calculation was repeated using tumor regions that had been eroded by one pixel using a 3×3 disk-shaped filter to investigate the robustness of this method and the influence of the tumor contours on the acquired results. DCE-CT parameters were also calculated using skeletonized regions to investigate whether similar results would be obtained using a smaller subset of the tumor pixels for the calculation of parameter values. Skeletonization is a morphological operation that iteratively deletes successive layers of pixels on the boundary of a region until a thin arc, representative of the original shape, remains [130]; in this case, DCE-CT parameters were calculated as the mean of pixels in the parameter map lying along the skeletonized region. Should this method obtain similar results as the full volumetric segmentation of dynamically imaged tumor, that could provide the basis for a more time-efficient sampling of tumor in practice and future studies. Figure 2.6 shows examples of original and skeletonized tumor regions from the dynamic section of a patient in this study.

2.6 Statistical Analysis

Statistical comparisons were made using Revolution R Enterprise 6.1.0 (Revolution Analytics, Mountain View, CA) running R 2.14.2. The Shapiro-Wilk test revealed that relative changes in DCE-CT parameter values between scans did not follow normal distributions; therefore, the two-sided Wilcoxon rank-sum test was used to test the null hypothesis that distributions of relative changes in BF and BV were identical between patients on treatment and patients on observation. Nonparametric 95% confidence intervals (CIs) for the median differences in relative change in BF and BV between scans when comparing a patient on treatment with a patient on observation were obtained through a bootstrapping method, for which 5000 bootstrapping samples were generated with replacement from the two patient groups [131]. Spearman's rank correlation coefficient ρ was calculated to assess the correlation between relative change in DCE-CT parameter values and relative change in tumor bulk between scans. The significance level was set at $\alpha = 0.05$ for all statistical tests, with the Bonferroni-Holm method used to correct for multiple comparisons [132]. Since $n = 14$ statistical tests were made, the nominal significance level of the test that achieved the lowest p-value was set to $\alpha' = \frac{\alpha}{n} = 0.004$, with subsequently higher nominal significance levels $\frac{\alpha}{n-1}, \frac{\alpha}{n-2}, \ldots$ for other statistical comparisons, in the order of increasing p-value.

The Bland Altman method was used to estimate agreement between the original tumor regions, eroded tumor regions, and skeletonized tumor regions for the calculation of mean tumor BF and BV values across all patients. DCE-CT parameter values calculated on scans of the same patient were treated as independent observations for the analysis. Differences between the methods did not follow a normal distribution in all cases; however, as the nonparametric approach of calculating limits of agreement is generally less reliable than the limits of agreement found using normality theory (especially for small samples), the 95% limits of agreement were estimated in the standard manner as the mean of the differences $\pm 1.96 \times s$, where s is the standard deviation of the differences between the two methods being compared. The standard error of the 95% limits of agreement was estimated as $\sqrt{3s^2/n}$,

where n is the number of scans, following the method of Bland and Altman [133, 134].

2.7 Results

2.7.1 Tumor Volume

The median volume of mesothelioma tumor that was dynamically imaged and used for DCE-CT parameter calculation across all scans (including both scans of each patient) was 41.6 cm^3 (range: 8.6–149.9 cm^3). This corresponded to a median fraction of 9% (range: 4%–75%) of the total volume of mesothelioma tumor within each patient's thorax, as estimated using an in-house semi-automated tumor segmentation method [53].

2.7.2 DCE-CT Parameters

Figure 2.7 shows mean DCE-CT parameter values for scan 2 as a function of scan 1 for both patient cohorts. Table 2.2 shows the median relative change in BF and BV between the two scans for patients on treatment and for patients on observation. The distributions of relative change in BF and BV between scans in the two patient groups were not found to be significantly different (p-value 0.19 and 0.06 for BF and BV, respectively). The 95% bootstrap-estimated CI for the median difference between the patient groups was (-19.6%, 4.3%) for BF and (-23.3%, 3.5%) for BV.

Figure 2.8 shows relative changes in BF and BV as a function of relative change in tumor thickness as measured by modified RECIST. Table 2.3 presents rank correlation statistics comparing mean relative change in BF and BV between scans to relative change in tumor thickness between scans as measured using modified RECIST; p-values in Table 2.3 are based on a null hypothesis of $\rho = 0$. No significant correlation was found between change in either BF or BV and change in tumor bulk.

Table 2.2: Median change in tumor DCE-CT parameters between scans for the two patient groups. p-values were calculated using a two-sided Wilcoxon rank-sum test. 95% CIs estimated using a bootstrap method with 5000 iterations.

DCE-CT Parameter	Patient Group	Median Change (Range)	p-value (95% CI of Median Difference Between Groups)
BF	Treatment	-16.1% (-22.2%–8.4%)	0.19 (-19.6%, 4.3%)
	Observation	-1.9% (-6.7%–0.1%)	
BV	Treatment	-9.4% (-18.6%–11.4%)	0.06 (-23.3%, 3.5%)
	Observation	5.8% (-1.5%–14.5%)	

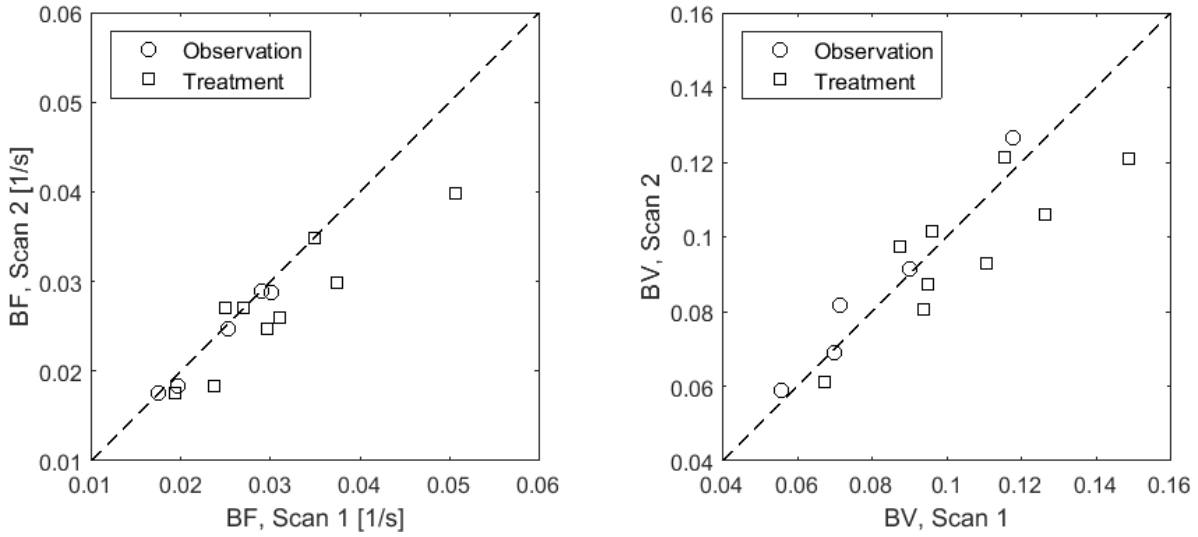

Figure 2.7: Mean DCE-CT parameter values for both DCE-CT scans for patients on observation and patients on treatment: (left) BF, (right) BV. The dashed line is the line of equality.

2.7.3 Assessment of Tumor Region Dependence

Table 2.4 shows the median change in BF and BV between scans when using the original tumor regions (compare with Table 2.2), the tumor regions eroded by one pixel, and skeletonized tumor regions to calculate mean DCE-CT parameter values. Figures 2.9 and 2.10 show Bland Altman plots comparing BF and BV values, respectively, obtained using the three variations of the tumor regions investigated in this study. The mean difference in BF acquired with the original regions and the eroded regions was -0.0002 s^{-1} with 95% limits of agreement [-0.0011 s^{-1}, 0.0008 s^{-1}] (95% CIs: [-0.0014 s^{-1}, -0.0008 s^{-1}] and [0.0005 s^{-1},

Table 2.3: Spearman rank correlation coefficients ρ and the corresponding p-values for correlation between relative changes in DCE-CT parameters and relative change in tumor thickness as measured using modified RECIST. The p-values were calculated for a null hypothesis of $\rho = 0$.

DCE-CT Parameter	$\rho_{\text{All Patients}}$	$\rho_{\text{Treatment}}$	$\rho_{\text{Observation}}$
BF	-0.165 ($p = 0.57$)	-0.05 ($p = 0.91$)	0.1 ($p = 0.95$)
BV	-0.437 ($p = 0.12$)	-0.433 ($p = 0.25$)	-0.5 ($p = 0.45$)

Table 2.4: Median change in tumor DCE-CT parameters between scans for the two patient groups using the original tumor regions, the tumor regions eroded by one pixel, and skeletonized tumor regions. p-values were calculated using a two-sided Wilcoxon rank-sum test.

Tumor Regions	DCE-CT Parameter	Patient Group	Median Change	p-value
Original	BF	Treatment	-16.1%	0.19
		Observation	-1.9%	
	BV	Treatment	-9.4%	0.06
		Observation	5.8%	
Eroded	BF	Treatment	-14.9%	0.19
		Observation	0.8%	
	BV	Treatment	-13.8%	0.04
		Observation	5.7%	
Skeletonized	BF	Treatment	-15.5%	0.06
		Observation	-5.0%	
	BV	Treatment	-8.8%	0.11
		Observation	2.4%	

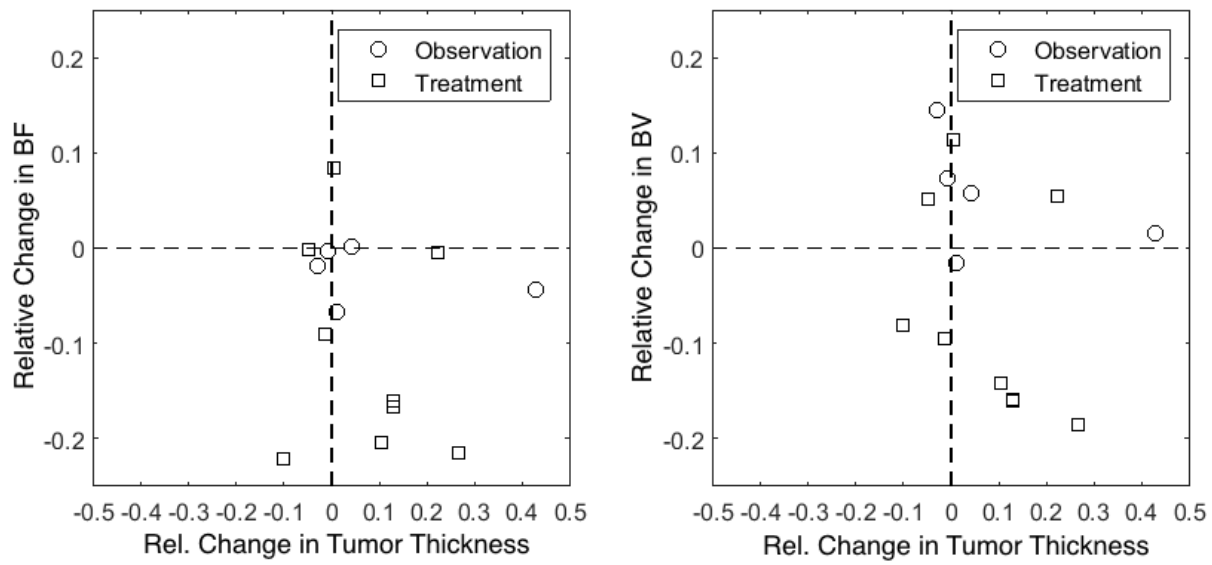

Figure 2.8: Relative change in mean DCE-CT parameter values in patients on observation and patients on treatment as a function of relative change in tumor thickness as measured by modified RECIST: (left) BF, (right) BV. The dashed lines represent no change between scans. Note that the scale of the x-axis differs by a factor of two from that of the y-axis.

0.0011 s^{-1}] for the lower and upper limit, respectively). The mean difference in BF acquired with the original regions and the skeletonized regions was -0.0002 s^{-1} with 95% limits of agreement [-0.0023 s^{-1}, 0.0020 s^{-1}] (95% CIs: [-0.0030 s^{-1}, -0.0016 s^{-1}] and [0.0013 s^{-1}, 0.0027 s^{-1}] for the lower and upper limit, respectively). The mean difference in BV acquired with the original regions and the eroded regions was -0.001 with 95% limits of agreement [-0.007, 0.005] (95% CIs: [-0.009, -0.005] and [0.003, 0.007] for the lower and upper limit, respectively). The mean difference in BV acquired with the original regions and the skeletonized regions was -0.002 with 95% limits of agreement [-0.016, 0.011] (95% CIs: [-0.020, -0.011] and [0.007, 0.015] for the lower and upper limit, respectively).

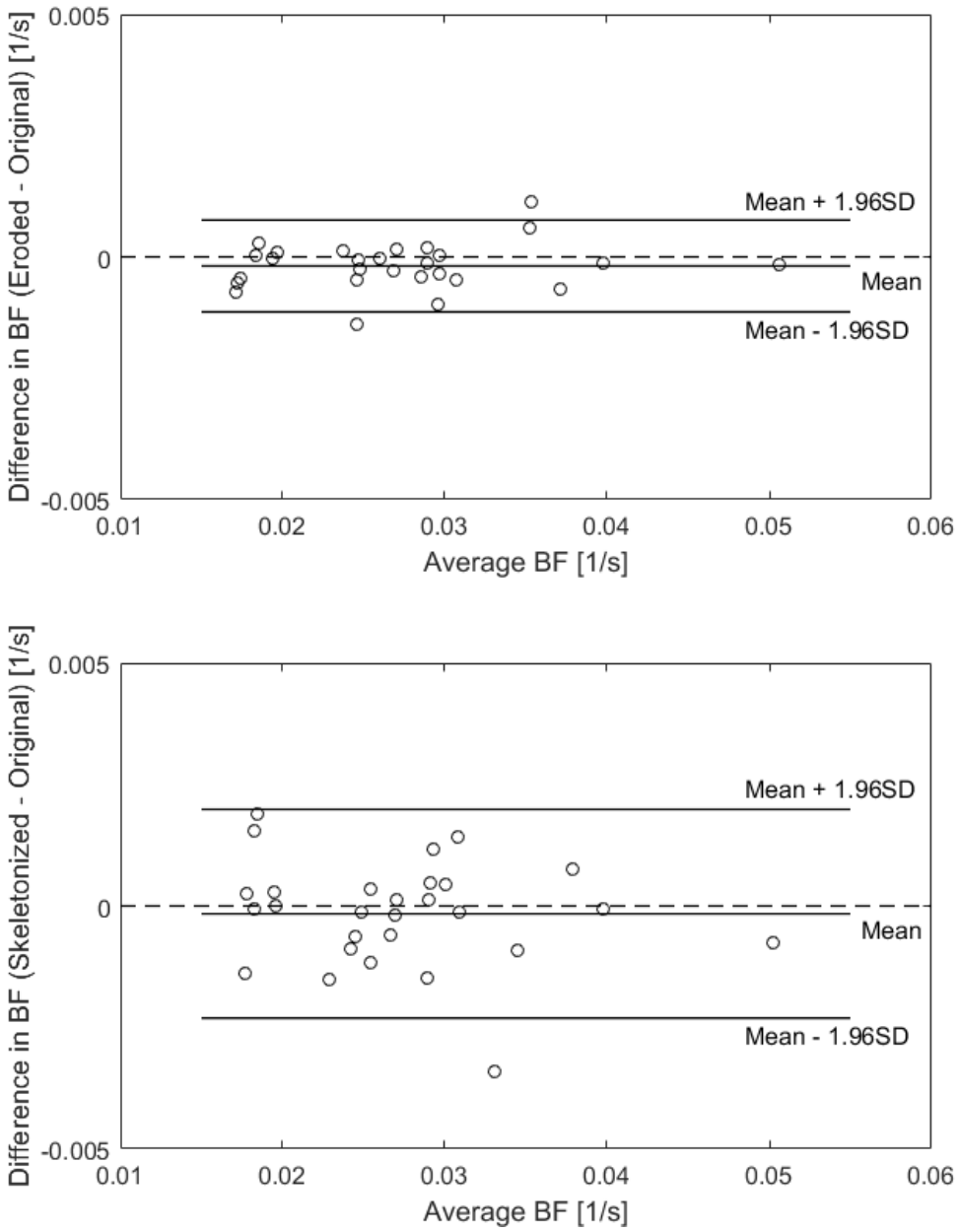

Figure 2.9: Bland Altman plots comparing BF values acquired (top) using the original tumor regions and the tumor regions eroded by one pixel and (bottom) using the original tumor regions and the skeletonized tumor regions. The mean difference and 95% limits of agreement are plotted as solid lines. The dashed line represents no difference in BF between variations of the tumor regions.

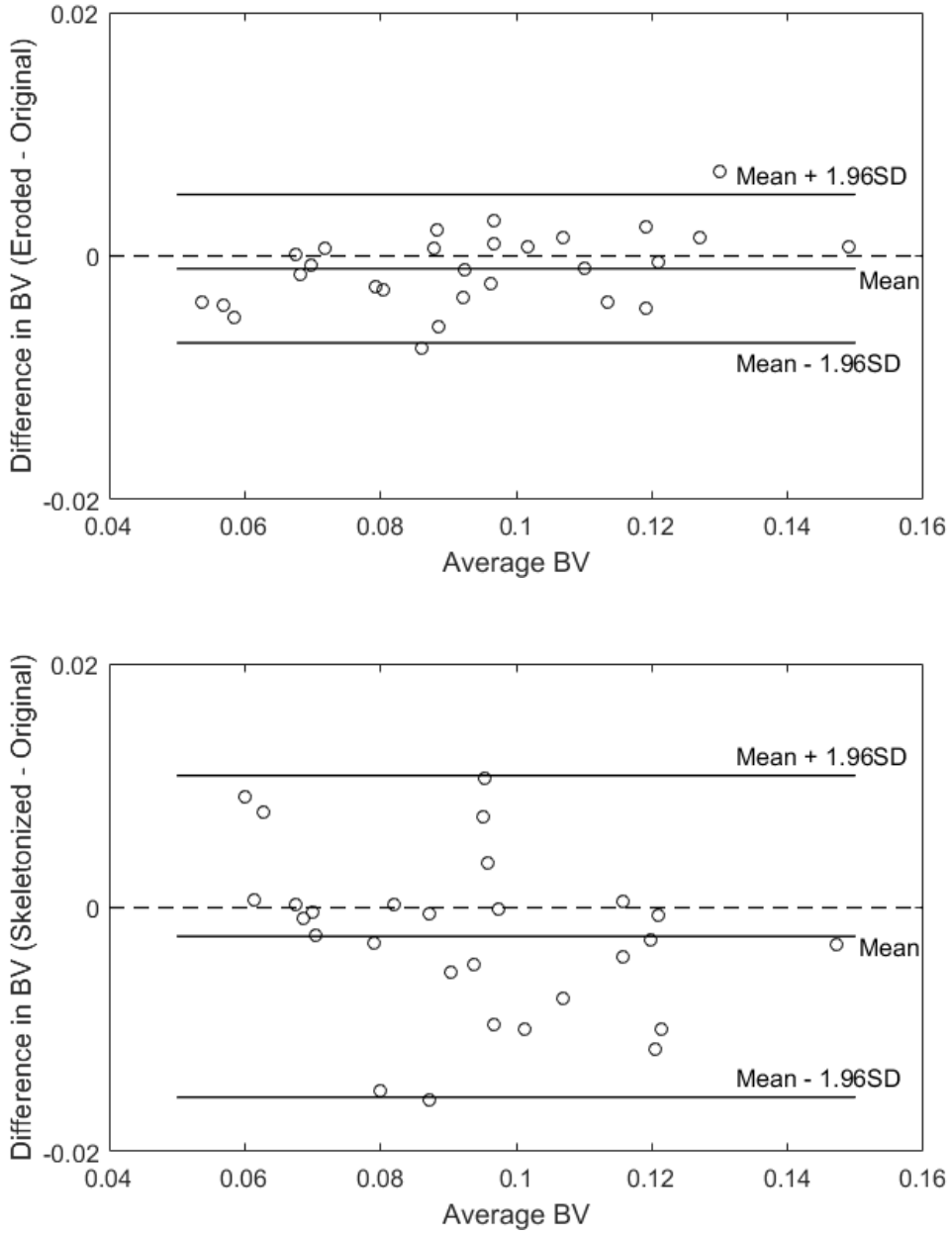

Figure 2.10: Bland Altman plots comparing BV values acquired (top) using the original tumor regions and the tumor regions eroded by one pixel and (bottom) using the original tumor regions and the skeletonized tumor regions. The mean difference and 95% limits of agreement are plotted as solid lines. The dashed line represents no difference in BV between variations of the tumor regions.

2.8 Discussion

Compared with patients on observation, patients on treatment in this study showed an overall decrease in BF and BV between scans. Although these differences did not reach statistical significance, the observed trends suggest the potential utility of DCE-CT parameter analysis in the assessment of pharmacokinetic endpoints for mesothelioma patients. The decrease in both BF and BV between scans for patients on treatment is consistent with the expectation that chemotherapeutic agents disturb tumor function. No significant rank correlation was found between relative change in DCE-CT parameters between the two scans of each patient and relative change in tumor thickness between the two scans as measured according to modified RECIST. This result could be due to the greater sensitivity of DCE-CT parameters to the altering effects of chemotherapeutic agents on tumor function relative to traditional response measures, which rely solely on disease morphology; additional studies will be required to further evaluate this effect.

Differences in DCE-CT parameter values (both BF and BV) acquired when using the original tumor regions and regions eroded by one pixel were small, and similar trends were observed in change in BF and BV between scans across the two patient groups for both original and eroded tumor contours (see Table 2.4). Erosion of tumor regions can be expected to reduce the probability of erronous inclusion of surrounding non-tumorous tissue in the calculation of mean DCE-CT parameter values. The observed agreement in mean DCE-CT parameter values across non-eroded and eroded tumor regions suggests that the method used for BF and BV calculation in this study is stable with respect to minor modifications of the regions.

DCE-CT parameter values calculated using skeletonized tumor regions varied to a greater extent than eroded regions when compared with the original tumor regions (see Figs. 2.9 and 2.10). Skeletonization of tumor regions was used in this study to investigate whether coarse sampling of the tumor volume could be used as a substitute for time-consuming volumetric segmentation of the tumor across the dynamic sections in the calculation of

DCE-CT parameters. Similar trends were observed in the change in BF and BV across patient groups with the skeletonized tumor regions as were observed with the original tumor regions (see Table 2.4). The small patient cohort in this study did not allow any statistically meaningful conclusions regarding this aspect of the study; however, the results indicate the validity of investigating the skeletonization approach for future studies of larger populations.

Only one other study has applied DCE-CT to the assessment of mesothelioma tumor response; Meijerink et al. [135] included two mesothelioma patients within their cohort of 16 patients, and BF was found to have decreased substantially for both patients following antiangiogenic treatment. The patient cohort of the current study included two patients on antiangiogenic treatment (ARQ 197 and GDC-0980); the relative change in BF between scans for these two patients was found to be -0.4% and -21.4%, compared with -47.1% and -18.6% for the patients in the study by Meijerink et al.

The unique morphology and growth pattern of mesothelioma affect the analysis of DCE-CT datasets. The axial and lateral extent of mesothelioma tumor can differ greatly from other cancers of the thorax to which DCE-CT analysis has previously been applied (e.g., non-small cell lung cancer). At the time of this study, commercial software packages for analyzing DCE-CT images were limited with regard to analysis of irregularly shaped volumetric regions that span multiple axial sections. This shortcoming necessitated the development of in-house software specifically designed for analysis of such regions, which allowed for DCE-CT analysis across multiple axial sections and with irregularly shaped tumor regions.

This study had several limitations. The patient cohort for the study was small, and the heterogeneity of the treatment cohort in terms of treatment regimens further reduced the statistical power of the analysis. Furthermore, any conclusions drawn from the differing trends observed between the two patient groups in this study rely on the assumption that tumor in patients on observation developed between scans "unhindered" and that the observed hemodynamics were characteristic of that state, and that trends observed for patients on treatment were secondary to treatment.

A future DCE-CT study of mesothelioma patients could be designed to include only patients undergoing the same treatment, which would allow for correlation of hemodynamic change with patient survival. The second dynamic imaging phase of the DCE-CT protocol for this study was deemed unusable for DCE-CT parameter comparison between patients due to variability in the starting time of this phase between scans. For example, another hemodynamic parameter, the mean transit time (a measure of tumor contrast washout), requires imaging at the latter end of the DCE-CT protocol and so could not be calculated in this study. The imaging protocol for future studies could be modified to standardize the start time of the second dynamic phase to ensure that sufficient wash-out has occurred and other parameters can be calculated and incorporated in the analysis.

2.9 Conclusion

In this study, the standard CT imaging protocol for mesothelioma patients was modified to include two phases of DCE-CT imaging. The final patient cohort consisted of 14 patients, of whom nine patients were on chemotherapeutic treatment during the course of the study and five patients were on observation. Each patient underwent two DCE-CT scans. No statistically significant differences in the relative change in the DCE-CT parameters BF and BV were found between the two patient groups, although both parameters showed different trends across the two patient groups, thus indicating the potential sensitivity of DCE-CT to assess pharmacodynamic endpoints in the treatment of mesothelioma. No significant correlations were found between relative changes in DCE-CT parameters and changes in tumor thickness as measured according to modified RECIST. Further evaluation of the relationship between DCE-CT parameter change and tumor response to treatment or patient survival will require a larger and more consistent group of patients on unified treatment regimens.

CHAPTER 3

DEEP CONVOLUTIONAL NEURAL NETWORKS FOR MESOTHELIOMA TUMOR SEGMENTATION ON CT SCANS

3.1 Introduction

As outlined in Section 1.1.7, tumor volume has been shown to correlate with patient outcomes in mesothelioma, and it has been suggested that image-based tumor volume could be a more representative measure of tumor bulk for the staging of mesothelioma, assessment of tumor response to treatment and as a predictor of patient survival [50–55, 58]. The time-consuming nature of manual (or even semi-automated) volumetric segmentation of measurable mesothelioma tumor throughout the entire hemithorax currently precludes the use of image-based volume in clinical practice; an accurate automated computerized volumetric segmentation of mesothelioma tumor could streamline the acquisition of such measurements. One published study presented a step-wise method for the automated segmentation of mesothelioma tumor on CT scans [56]. Traditional step-wise segmentation methods carry the inherent vulnerability that if one stage of the process fails, the tumor segmentation will be unsuccessful. On the other hand, given a model architecture that is of sufficient learning capacity and is applicable to the segmentation task, a deep CNN trained on a sufficiently extensive and varied set of reference segmentations has the potential to bypass this limitation of traditional approaches.

In this study[1], we investigated the utility of deep CNNs based on the U-Net deep CNN architecture for the automated segmentation of mesothelioma tumor on CT scans. This task is challenging due to variability in disease presentation and the low contrast of mesothelioma tumor relative to surrounding soft-tissue structures [17, 18]. The ability of deep CNN architectures to effectively learn and combine local and global image features in their classification model could provide a key to the robust volumetric segmentation of mesothelioma tumor.

1. This study has been published in a peer-reviewed journal [136].

3.2 Scope of Study

Mesothelioma patients often present with pleural effusion and atelectatic lung adjacent to the pleural space, both of which have considerable overlap in HU values with mesothelioma tumor [17, 18]. Early-stage mesothelioma patients exhibit unilateral disease, although in later stages of the disease the tumor may extend to the contralateral pleura or invade nearby structures such as the mediastinum, peritoneum, and chest wall [26]. As an initial effort at implementing a deep CNN-based method for the segmentation of mesothelioma tumor, this study focused on identifying pleural thickening (which includes predominantly tumor, along with potential pleural effusion and pleural plaques) in patients with unilateral disease where the tumor had not invaded other organs or structures.

Two deep CNNs were trained separately in the left and right hemithoraces on the two-class problem of differentiating between pleural thickening and "background" pixels on axial CT sections. Results of the present deep CNN-based segmentation method were compared with (1) the output of a previously published automated mesothelioma tumor segmentation method and (2) manual tumor outlines constructed on two sets of CT sections not included in the training dataset; one set of scans had tumor outlines constructed by a group of three observers (two attending thoracic radiologists, one radiology resident), and the other set of scans had tumor outlines constructed by five attending thoracic radiologists [56].

3.3 Data Preprocessing

The patient couch and surrounding air was segmented out from all CT scans used for training, validation, and testing through an in-house thoracic segmentation method developed in Matlab (Mathworks Inc., Natick, MA). All CT sections used for training, validation, and testing were converted to unsigned 8-bit integer images with a linear scaling such that pixels outside the thorax and pixels of value equal to or below -1000 HU were assigned a value of 0, and pixels of value equal to or greater than 400 HU were assigned a value of 255. This

rescaling of pixel values was based on preliminary investigations on a subset of the training set that showed that capturing the structure of the lungs could be advantageous with respect to distinguishing between tumor pixels in the pleural space and soft-tissue pixels along the outside of the thorax.

3.4 Training Set

234 CT scans from 87 mesothelioma patients were retrospectively collected for training the networks of this study. These images were a subset of those analyzed in a previously published method on the use of disease volumes as a marker for patient response in mesothelioma [53]. Pleural thickening was outlined on the scans of the training set by an imaging scientist trained in thoracic anatomy using a semi-automated segmentation method. Of the 87 patients, 39 patients (103 scans) had disease in the left hemithorax and 48 patients (131 scans) had disease in the right hemithorax. Slice thickness varied across scans (see Table 3.1); to reduce the probability that scans containing a relatively large number of axial sections would overly influence the training process, only every other section was included in the training set for scans of slice thickness less than 2 mm and only every third section was included for scans of slice thickness less than 1 mm.

3.5 Test Sets

Two test sets of mesothelioma patient scans with radiologist-provided reference tumor segmentations were used for testing the deep CNNs trained in this study.

Test set 1 consisted of 61 axial CT sections from 16 patients with pathologically confirmed mesothelioma, with reference segmentations provided independently by two attending radiologists and one radiology resident (Observers A, B and C). These images were used in the analysis of a previously published method on the automated segmentation of mesothelioma tumor on CT scans (the "2011 Method") and provided a direct comparison with that

Table 3.1: Characteristics of 234 patient scans available for training of the deep CNNs.

Characteristic	Value
Disease laterality	
Left hemithorax	39 (103 scans)
Right hemithorax	48 (131 scans)
Median no. of segmented sections per scan	
Scans with left-sided disease	61 (range: 28–167) sections
Scans with right-sided disease	69 (range: 35–153) sections
Median slice thickness	
Scans with left-sided disease	5 (range: 0.625–10) mm
Scans with right-sided disease	2.5 (range: 0.625–7) mm
Median pixel spacing	
Scans with left-sided disease	0.703 (range: 0.582–0.871) mm
Scans with right-sided disease	0.703 (range: 0.543–0.836) mm

method [56]. Scans containing prominent calcifications or surgical intervention were excluded from this test set, as were CT sections for which no disease was present or the mean Dice similarity coefficient (DSC) value across all observers was less than or equal to 0.5 (thus reflecting complex disease with low observer agreement). The DSC is defined as a function of the area of intersection and the sum of the area of a pair of segmentations under comparison [137, 138]:

$$\mathrm{DSC} = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}, \tag{3.1}$$

where $|S_1|$ and $|S_2|$ represent the respective area of each segmentation and $|S_1 \cap S_2|$ represents the area of the intersection of the two segmentations. Furthermore, sections for which all three observers did not agree on the laterality of disease were excluded from the analysis of the present study due to the hemithorax-specific nature of the present deep CNN-based segmentation method. Of the 61 axial sections, 42 had right-hemithorax disease and 19 had left-hemithorax disease (see Table 3.2).

Test set 2 consisted of 70 axial CT sections from the baseline scans of 27 patients with pathologically confirmed mesothelioma, with reference segmentations provided independently by five attending thoracic radiologists (Observers 1, 2, 3, 4 and 5). These images

Table 3.2: Characteristics of CT sections used for testing the deep CNN-based segmentation method.

Characteristic	Value
Test set 1 ($n = 61$)	
Sections with left-sided disease	19 sections (31%)
Sections with right-sided disease	42 sections (69%)
Median slice thickness	1 (range: 1–2) mm
Median pixel spacing	0.709 (range: 0.629–0.861) mm
Test set 2 ($n = 70$)	
Sections with left-sided disease	21 sections (30%)
Sections with right-sided disease	49 sections (70%)
Median slice thickness	3 (range: 3–5) mm
Median pixel spacing	0.734 (range: 0.535–0.883) mm

were used in a previously published study on observer variability in mesothelioma tumor area measurements [53]. Sections for which all observers did not agree on the presence or laterality of disease, and for which the mean DSC value over all observers was less than or equal to 0.5, were excluded from the present analysis. Of the 70 axial sections, 49 had right-hemithorax disease and 21 had left-hemithorax disease (see Table 3.2).

3.6 Deep CNN Architecture

The U-Net deep CNN architecture presented by Ronneberger et al. [116] was used for the classification of pixels as pleural thickening or background on axial CT sections. Fig. 3.1 shows the architecture of the deep CNN used in the present study. The network accepted as input a 512×512 image matrix, consisted of a contracting path and an expansive path, and produced a tumor segmentation mask of the same size as the input. At each level of the contracting path, two 3×3 convolutional layers were applied to the input matrix or the matrix output by the previous level. Convolutional layers in the down- and upsampling paths of the network were followed by a ReLU activation function [139]. Following the convolutional layers at each level, the downsampling of the matrix was achieved through a 2×2 max pooling operation with stride 2. As described in the original U-Net paper, the number of

feature maps (i.e., the number of trained filters) was doubled at each downsampling step, starting with 64 maps at the input level of the network. At levels for which the downsampled matrix had reached a size of 64×64 pixels or smaller, the second ReLU was additionally followed by a Dropout layer at which individual neurons were deactivated at random during training with a probability of 0.5 to prevent overfitting [109].

The upsampling path of the network was initiated once the downsampled feature matrix reached dimensions of 32×32, with 1024 feature maps at that level. At each level of the upsampling path, a two-dimensional upsampling operation using nearest-neighbor interpolation was applied to the feature matrix, and the resulting feature map was concatenated with the feature map from the corresponding level of the downsampling path. Two 3×3 convolutional layers were applied to the resulting feature matrix at each level of the upsampling path. When the feature matrix had reached dimensions of 512×512 pixels, the max pool operation was replaced by a 3×3 convolution with two output feature maps. Finally, at the output of the network, a pixel-wise probability matrix of dimensions 512×512 was generated through a 1×1 convolutional layer followed by a sigmoid activation function. Binary tumor segmentation masks were generated by thresholding the probability matrix at the value of 0.5, which was chosen prior to the start of the study.

Network loss during training was calculated as the cross-entropy L averaged over all pixels of each deep CNN-predicted segmentation and the corresponding reference segmentation:

$$L(t_i, p_i) = -\left(t_i \log\left(p_i\right) + \left(1 - t_i\right) \log\left(1 - p_i\right)\right), \tag{3.2}$$

where t_i is an indicator variable taking the value 1 if the reference classification of pixel i is tumor and 0 otherwise, and p_i is the (continuous) deep CNN-predicted probability that pixel i is tumor ($p_i = 1$) or background ($p_i = 0$).

The Adam method was used to optimize the network during training using a learning rate of 10^{-4}, chosen prior to the study [140]. The deep CNN architecture was implemented

using the Keras and Tensorflow deep learning frameworks [141]. Experiments were run using online learning (i.e., a batch size of 1) on a scientific computing cluster at the University of Chicago using Nvidia GeForce GTX Titan and Nvidia Tesla K20c Kepler-class graphics processing units (GPUs; Nvidia, Santa Clara, CA).

DSC values acquired on the test sets were calculated according to Eq. (4.2) using binary segmentation masks after thresholding the continuous tumor probability matrices output by the deep CNNs. DSC values acquired on the training and validation sets during training were approximated using the continuous probability values output by the CNNs, due to computational considerations of running deep CNN code with the GPU-based Tensorflow architecture, which precludes efficient thresholding of the output matrices during training to generate binary segmentation masks. For the predicted output of a given image in the training set, we have:

$$\mathrm{DSC}_{\mathrm{Cont.}} = \frac{1 + 2 \times \sum_{i,j} p_{i,j} p'_{i,j}}{1 + \sum_{i,j} p_{i,j} + \sum_{i,j} p'_{i,j}} \quad , \tag{3.3}$$

where $p_{i,j}$ is the class (i.e., tumor or background) of the pixel in row i and column j of the reference tumor segmentation and $p'_{i,j}$ is the predicted probability output by the CNN. The addition of 1 in the numerator and denominator of Eq. (3.3) prevents numerical instability for images for which the predicted tumor probability is low across the whole image. Values calculated on the training and validation sets according to Eq. (3.3) will be labeled as $\mathrm{DSC}_{\mathrm{Cont.}}$ in the following sections.

3.7 Experiments

Deep CNNs were trained separately on the sections and reference segmentations of mesothelioma patients with visible disease in the left and right hemithorax. In cases for which more than two scans were available for a single patient, only the earliest and last available CT scans were used for training of the networks to reduce the influence of individual patients

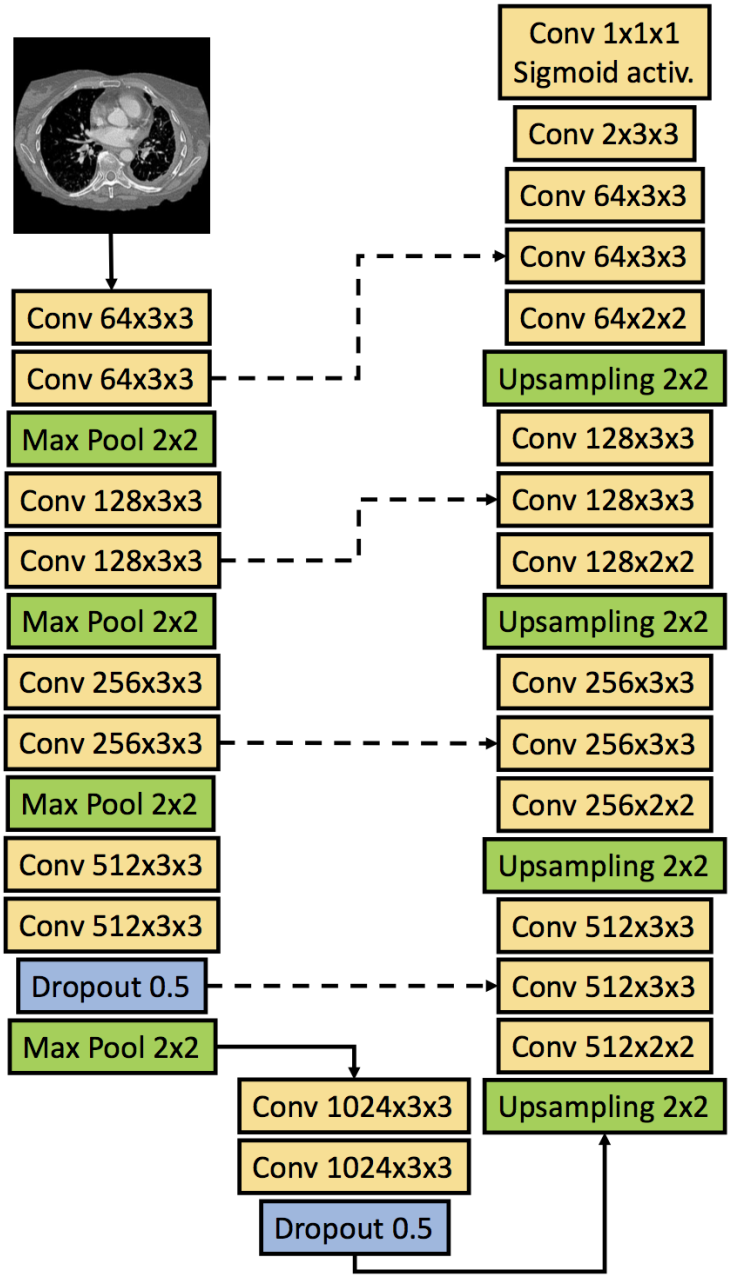

Figure 3.1: Architecture of the U-Net deep CNN for mesothelioma tumor segmentation. The deep CNN takes as input a 512×512 image matrix. Solid arrows indicate the flow of the input matrix through the network, and dashed lines indicate merging of information through concatenation of feature maps. Convolutional layers are labelled by the number of filters (equivalently, the number of feature maps) output by the layer, times the height and width of the convolution window. Individual neurons were "dropped" at a probability of 0.5 in the two dropout layers of the network. All convolutional layers used the ReLU activation function, except where noted. Upsampling was acquired through nearest neighbor interpolation.

on the training process. We randomly selected eight patients from the training set of each hemithorax classifier for inclusion in a validation set to evaluate the level of overfitting of the deep CNNs to the training sets. For this study, the selection of eight patients out of the available 39 and 48 patients (for the left and right hemithorax, respectively) for validation during training for each hemithorax was done at random; no matching of scan characteristics (e.g., tumor volume) was done during this selection process. Eight patients was assumed to be a sufficient number of patients to obtain a representative set of scans for validation in a patient group of this size without reducing the segmentation performance due to a reduction in the number of scans available for training the networks. A single scan was randomly selected out of all available scans for each of these eight patients (sixteen patients in total across both hemithoraces) for use in the validation sets. During training, the CNNs were applied to these validation sets after each training epoch (i.e., iteration over the training set). To evaluate the variance of the segmentation method performance, this process of validation set extraction was repeated two more times for each hemithorax, without replacement. Each of these pairs of training and validation sets were used to train a deep CNN. Table 3.3 shows the number of patients, scans, and sections for each of the training and validation sets of this study. Only the deep CNN trained using the first such selected validation set (i.e., "Training set 1" in Table 3.3) for each hemithorax was subsequently applied to the test sets; the corresponding training sets consisted of 4259 and 6192 axial sections in the left and right hemithorax, respectively.

Data augmentation is a technique in which deformations are applied to the images of the training set to increase the amount of data available for training and to improve CNN generalizability to other datasets [142]. The use of data augmentation was investigated in this study for the task of mesothelioma segmentation on axial CT sections. Minimal augmentation was applied to the set of axial sections used for training due to the inherent asymmetry of the anatomy of the chest. For this purpose, before each training iteration of the network, a random rotation in the range [-5°, +5°] and a random scaling in the

Table 3.3: Division of training and validation sets of each hemithorax. No scans or sections of the same patient formed a part of both the training and validation set of a given training-validation split.

Disease Laterality	Set	No. of Patients	No. of CT Sections
Left hemithorax	Training set 1	31 (54 scans)	4259
	Training set 2	31 (56 scans)	4354
	Training set 3	31 (56 scans)	4459
	Validation set 1	8 (8 scans)	652
	Validation set 2	8 (8 scans)	655
	Validation set 3	8 (8 scans)	643
Right hemithorax	Training set 1	40 (76 scans)	6192
	Training set 2	40 (75 scans)	6441
	Training set 3	40 (74 scans)	6391
	Validation set 1	8 (8 scans)	797
	Validation set 2	8 (8 scans)	595
	Validation set 3	8 (8 scans)	586

range [0.95, 1.05] were independently determined and applied to each of the sections used for training. The ranges of the rotation and scaling were determined prior to training by visualizing different rotation angles and scaling values on a subset of the training set.

The average binary cross-entropy L of Eq. 3.2 and the average $\mathrm{DSC}_{\mathrm{Cont.}}$ computed from the validation sets were used in each hemithorax to select the optimal deep CNN to apply to the test sets, with the objectives of (1) minimizing the average L on the initial validation set, (2) maximizing the average $\mathrm{DSC}_{\mathrm{Cont.}}$ on the initial validation set and (3) minimizing the variance in L and $\mathrm{DSC}_{\mathrm{Cont.}}$ across the three validation sets.

3.8 Statistical Analysis

Visual inspection revealed that the DSC values obtained when comparing the segmentations of the two computerized methods to the observer reference segmentations on test set 1 did not follow normal distributions. Therefore, the two-sided Wilcoxon signed-rank test was used to test the null hypothesis that the distributions of DSC values were identical for the present deep CNN-based method and the 2011 Method when compared with reference segmentations

by each of the three observers on test set 1. The Bonferroni correction was applied to the significance level of all statistical tests to account for the number of comparisons; since $m = 3$ statistical tests were made, the nominal significance level of individual comparisons was adjusted to $\alpha' = \alpha/m = 0.05/3 = 0.017$ to ensure that the family-wise error rate remained below or equal to $\alpha = 0.05$. Statistical comparisons were made using Matlab.

Bland-Altman plots were used to evaluate agreement between (1) the tumor area segmented by the 2011 Method and the average tumor area segmented by the three observers on test set 1 (results for the 2011 Method were not available for test set 2) and (2) between the tumor area segmented by the present deep CNN-based method and the average tumor area segmented by each set of observers on the two respective test sets. Absolute differences between the segmented area of the computerized methods and the average observer-segmented area were found to have a positive correlation with the average segmented tumor area of the segmentation approaches being compared, violating the normality assumption for calculation of 95% limits of agreement according to the Bland-Altman method [133]. Therefore, the 95% limits of agreement were estimated using relative differences in segmented area as $d \pm 1.96s$, where d is the mean and s is the standard deviation of the relative differences between the two segmentation approaches being compared (i.e., computerized and manual). The standard error (SE) of d was estimated as $\sqrt{(s^2/n)}$ and the SE of the 95% limits of agreement was estimated as $\sqrt{(3s^2/n)}$, where n is the number of segmented axial sections. 95% CIs for d and the 95% limits of agreement were found by adding and subtracting 1.96×SE from each value in question [134].

3.9 Results

3.9.1 Training

The binary cross-entropy loss L and $\mathrm{DSC}_{\mathrm{Cont.}}$ values on the training and validation sets, with and without data augmentation, are shown in Figure 3.2. The lines in Figure 3.2 represent

Table 3.4: Minimum binary cross-entropy loss L and the corresponding $\text{DSC}_{\text{Cont.}}$ value achieved on the initial validation set during training of the deep CNNs of each hemithorax and the range of L and $\text{DSC}_{\text{Cont.}}$ at the corresponding epochs across all three validation sets used to assess variance in segmentation performance during training. Values shown for networks trained with and without data augmentation. For the right-hemithorax deep CNN trained with augmentation, the network that achieved the second-lowest value of L on the initial validation set was selected for application to the test sets due to the narrower range of $\text{DSC}_{\text{Cont.}}$ values across the three validation sets at the corresponding epoch.

Hemithorax	Metric	Epoch	Value (Range)
Left (without augmentation)	Minimum L	15	0.021 (0.019–0.028)
	$\text{DSC}_{\text{Cont.}}$	15	0.752 (0.717–0.752)
Left (with augmentation)	Minimum L	19*	0.023 (0.019–0.024)
	$\text{DSC}_{\text{Cont.}}$	19*	0.746 (0.704–0.746)
Right (without augmentation)	Minimum L	11	0.030 (0.021–0.030)
	$\text{DSC}_{\text{Cont.}}$	11	0.741 (0.682–0.742)
Right (with augmentation)	Minimum L	9	0.024 (0.019–0.024)
	$\text{DSC}_{\text{Cont.}}$	9	0.732 (0.650–0.732)
	Second-lowest L	12*	0.026 (0.020–0.026)
	$\text{DSC}_{\text{Cont.}}$	12*	0.733 (0.677–0.733)

*Deep CNNs selected for application to the test sets

L and $\text{DSC}_{\text{Cont.}}$ values on the training and validation sets that were used to select optimal CNNs for application to the test sets; shaded areas indicate the range of L and $\text{DSC}_{\text{Cont.}}$ values over all three pairs of training and validation sets. Table 3.4 lists the minimum loss L achieved on the initial validation set for each hemithorax and the corresponding $\text{DSC}_{\text{Cont.}}$ value on the initial validation set, and ranges of L and $\text{DSC}_{\text{Cont.}}$ values across the three validation sets at the corresponding epoch, with and without data augmentation. Based on the performance metric values achieved on the first validation set of each hemithorax, for the left hemithorax, training epoch 19 was selected as the optimal deep CNN for application on the test sets; for the right hemithorax, epoch 12 was selected as the optimal deep CNN for application on the test sets. In both hemithoraces, the selected optimal deep CNNs were trained with data augmentation.

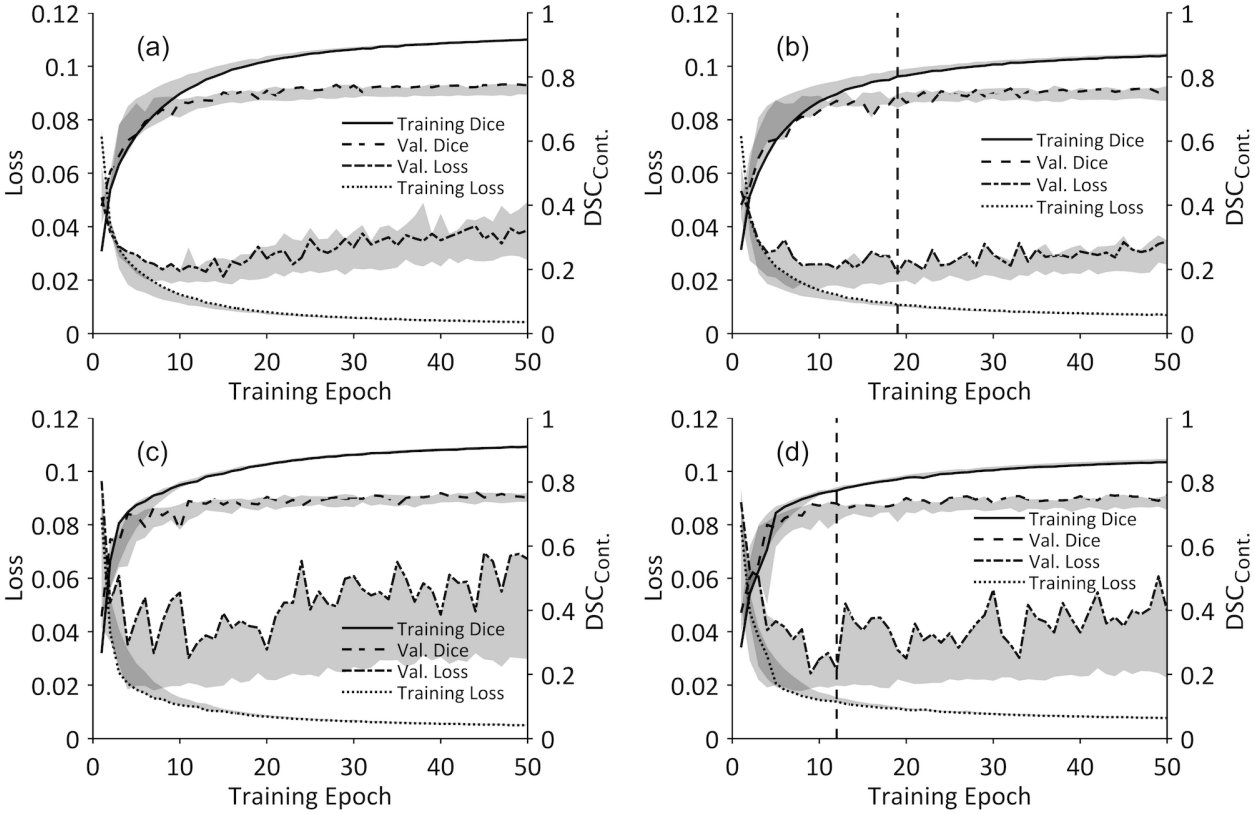

Figure 3.2: Binary cross-entropy loss and $\mathrm{DSC}_{\mathrm{Cont.}}$ on the training and validation sets during training of the left-hemithorax deep CNN ((a) without data augmentation and (b) with data augmentation) and of the right-hemithorax deep CNN ((c) without data augmentation and (d) with data augmentation). Lines indicate results on the initial pairs of training/validation sets. Shaded areas indicate the range of the loss and $\mathrm{DSC}_{\mathrm{Cont.}}$ across all three pairs of training/validation sets used to assess variance in segmentation performance during training. The vertical dashed lines (at epoch 19 in (b) and epoch 12 in (d)) indicate the training epochs after which the deep CNNs trained on the initial training/validation sets were applied to the test sets.

3.9.2 Test Set 1

Figure 3.3 shows boxplots of DSC values obtained when comparing the tumor segmentations of the present deep CNN method and the 2011 Method with the reference segmentations of the three observers on test set 1, and when comparing reference segmentations across observers on test set 1. The median DSC value for the deep CNN method was 0.776 (range: 0.314–0.938), 0.689 (range: 0.251–0.931) and 0.800 (range: 0.308–0.952) when compared with reference tumor contours by Observers A, B, and C, respectively. The median DSC value for the 2011 Method on the same CT sections was 0.720 (range: 0–0.938), 0.604 (range: 0–0.902), and 0.718 (range: 0–0.926) for Observers A, B, and C, respectively. Differences in the distributions of DSC values between the two automated segmentation methods on test set 1 were found to be statistically significant for all observers ($p < 0.0005$, $p < 0.00001$, and $p < 0.00001$ for Observers A, B, and C, respectively). The median DSC value for interobserver comparisons was 0.652 (range: 0.363–0.885), 0.648 (range: 0.293–0.902) and 0.814 (range: 0.461–0.937) when comparing Observers A and B, Observers B and C, and Observers A and C, respectively.

Figure 3.4 (a) shows a Bland-Altman plot of the relative differences in segmented tumor area by the deep CNN method and the average tumor area segmented by Observers A, B, and C on test set 1. The mean relative difference in segmented tumor area between the deep CNN method and the average observer-segmented area was -0.2% (95% CI: [-8.8%, 8.5%]) with 95% limits of agreement [-66.4%, 66.1%] (95% CIs: [-81.4%, -51.4%] and [51.1%, 81.1%] for the lower and upper limit, respectively).

Figure 3.4 (b) shows a Bland-Altman plot of the relative differences in segmented tumor area by the 2011 Method and the average tumor area segmented by the three observers on test set 1. The mean relative difference in segmented tumor area between the 2011 Method and the average observer-segmented area was 10.3% (95% CI: [-3.5%, 24.1%]) with 95% limits of agreement [-95.4%, 115.9%] (95% CIs: [-119.3%, -71.5%] and [92.0%, 139.8%] for the lower and upper limit, respectively).

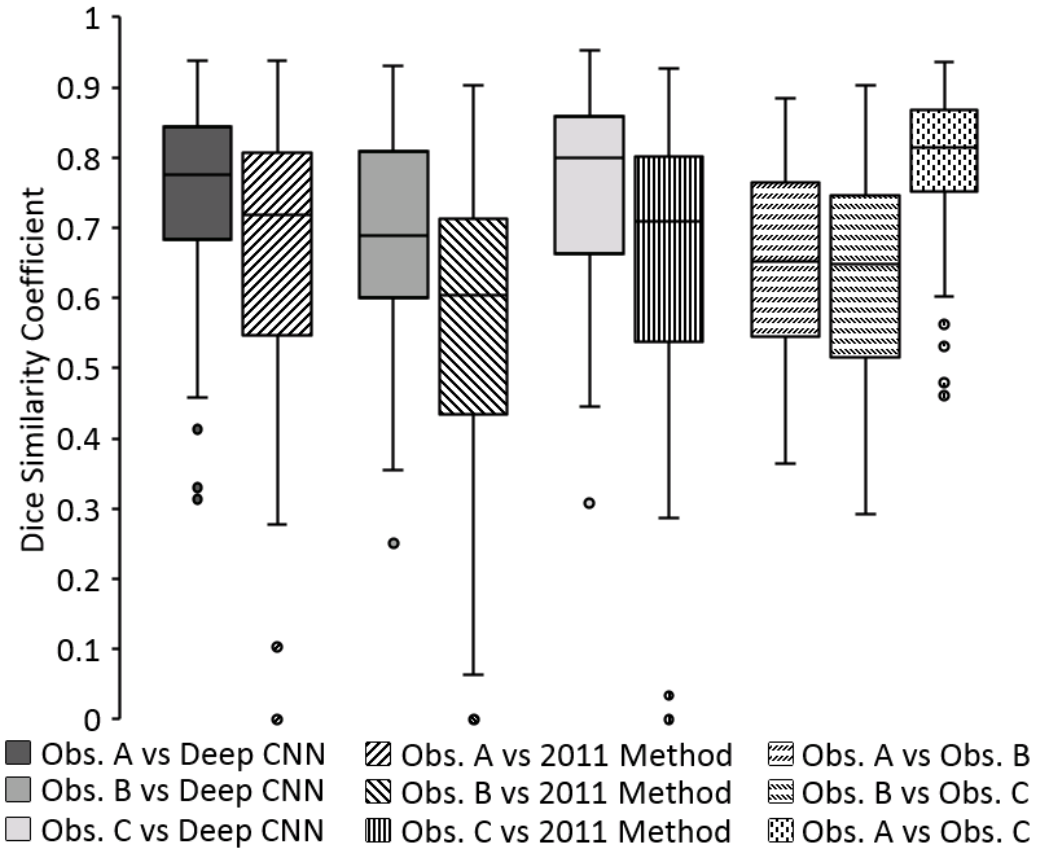

Figure 3.3: Boxplots showing DSC values obtained when comparing tumor segmentations by the present deep CNN-based method and the 2011 Method with reference segmentations of all three observers on test set 1 and when comparing reference segmentations across observers on test set 1. Horizontal lines inside boxes indicate the median value of each distribution.

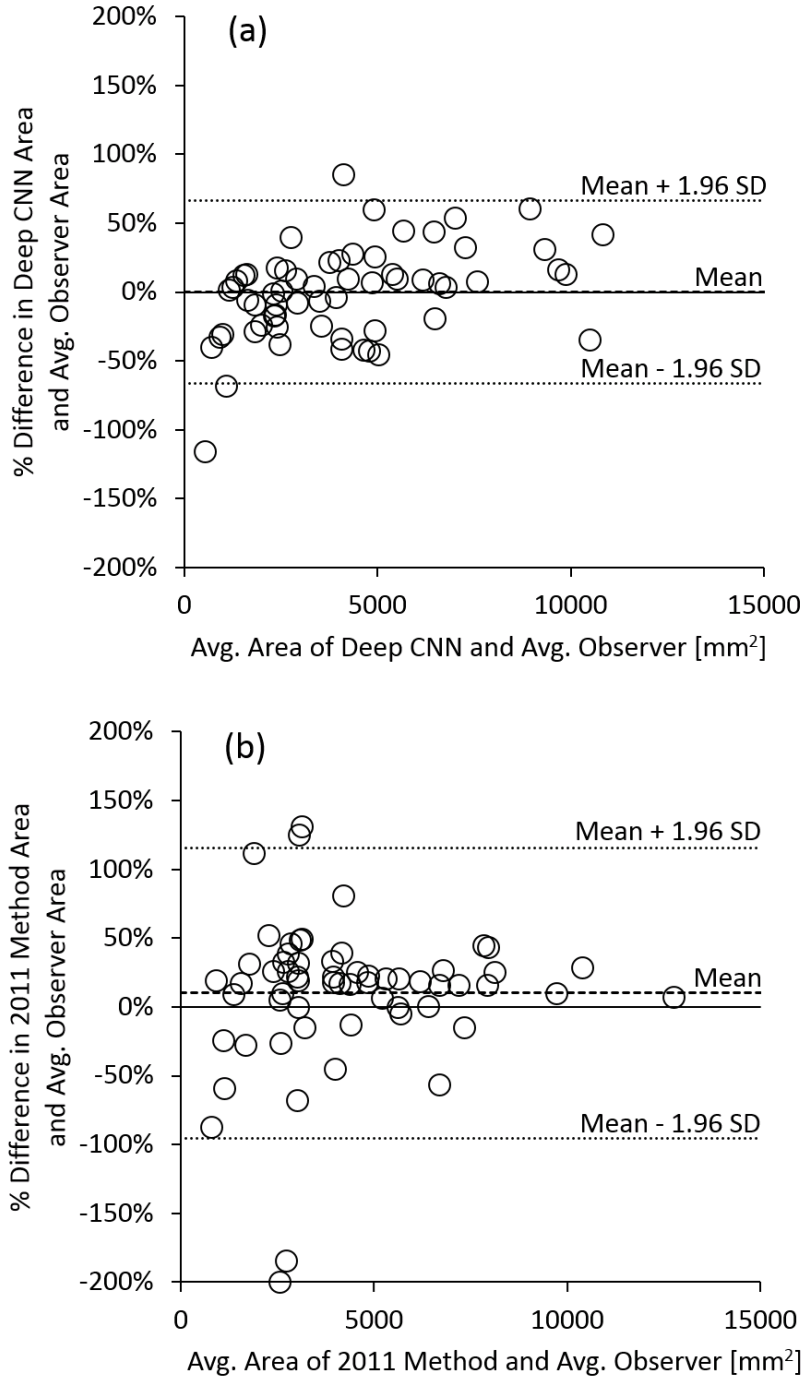

Figure 3.4: Bland-Altman plots showing (a) the relative differences between the segmented tumor area of the present deep CNN-based method and the average observer-segmented tumor area on test set 1 and (b) the relative differences between the segmented tumor area of the 2011 Method and the average observer-segmented tumor area on test set 1. Means of relative differences and 95% limits of agreement shown as dashed lines.

3.9.3 Test Set 2

Figure 3.5 shows boxplots of DSC values obtained when comparing tumor segmentations of the present deep CNN method with the reference segmentations of the five observers on test set 2 and when comparing reference segmentations across observers on test set 2. The median DSC value for the deep CNN method on test set 2 was 0.735 (range: 0.111–0.906), 0.662 (range: 0.086–0.879), 0.797 (range: 0.142–0.944), 0.747 (range: 0.108–0.919) and 0.755 (range: 0.148–0.921) when compared with reference tumor contours by Observers 1, 2, 3, 4, and 5, respectively. The median DSC value of the interobserver comparisons of the five observers on test set 2 ranged from 0.720 (range: 0.413–0.905) to 0.813 (range: 0.457–0.948).

Figure 3.6 shows a Bland-Altman plot of the relative differences in segmented tumor area by the deep CNN method and the average tumor area segmented by all five observers on test set 2. The mean relative difference in segmented tumor area by the deep CNN method and the average observer-segmented area was 19.5% (95% CI: [9.5%, 29.4%]) with 95% limits of agreement [-62.1%, 101.0%] (95% CIs: [-79.3%, -44.9%] and [83.8%, 118.3%] for the lower and upper limit, respectively). Seven out of the 15 sections that showed a greater than 29.4% relative difference (the upper limit of the 95% CI of the mean relative difference) in segmented tumor area exhibited pleural effusions that were classified as tumor by the deep CNN-based method and excluded from tumor segmentations by all five observers. Leaving these sections out of the calculation, the mean relative difference in deep CNN-predicted tumor area and the average tumor area segmented by the five observers on the remaining 63 sections of test set 2 was 8.7% (95% CI: [2.4%, 15.1%]) with 95% limits of agreement [-40.8%, 58.3%] (95% CIs: [-51.8%, -29.8%] and [47.2%, 69.3%] for the lower and upper limit, respectively).

Figure 3.7 shows the preprocessed CT sections, observer reference segmentations, and deep CNN tumor segmentations for three example CT sections selected at random from the bottom 10th percentile, the interquartile range, and the top 10th percentile of the average DSC value when comparing deep CNN tumor segmentations with observer reference segmentations across both test sets.

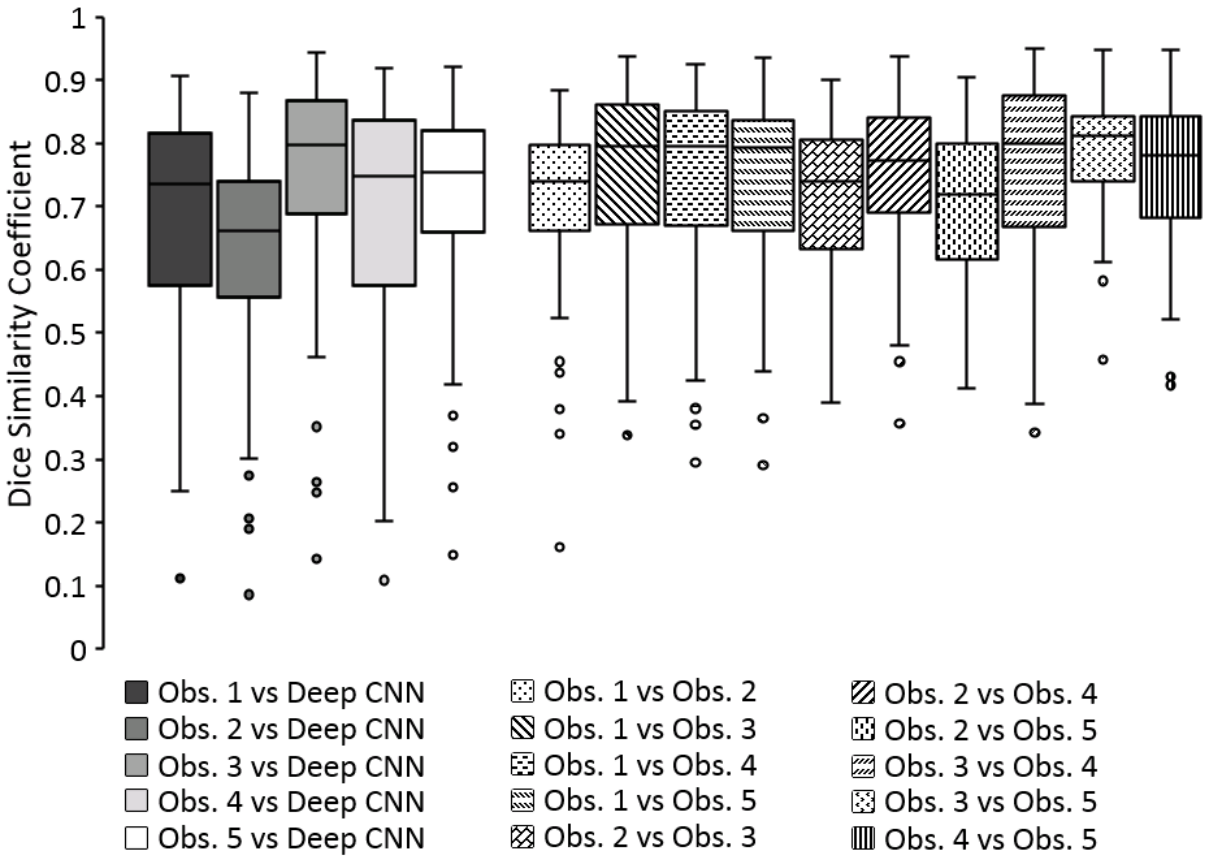

Figure 3.5: Boxplots showing the DSC values obtained when comparing the predicted tumor segmentations by the present deep CNN method with observer reference segmentations on test set 2 and when comparing reference segmentations across observers on test set 2. Horizontal lines inside the boxes indicate the median value of each distribution.

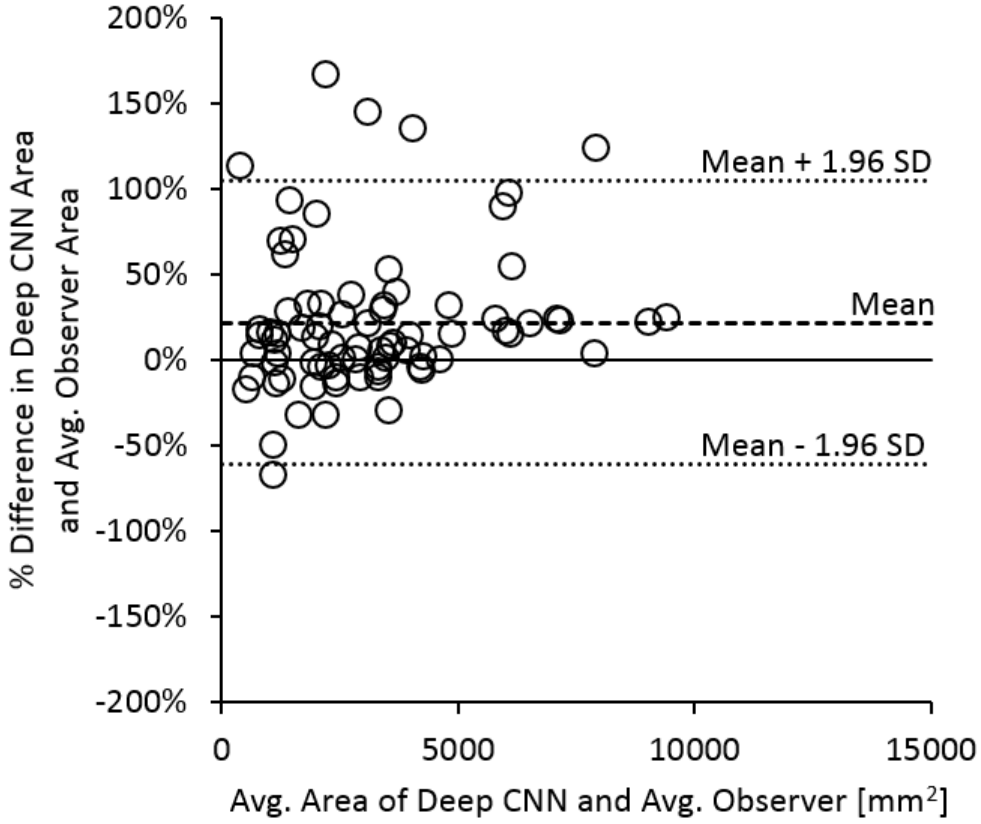

Figure 3.6: Bland-Altman plot showing the relative differences between the segmented tumor area of the present deep CNN-based method and the average observer-segmented tumor area on test set 2. Mean of relative differences and 95% limits of agreement shown as dashed lines.

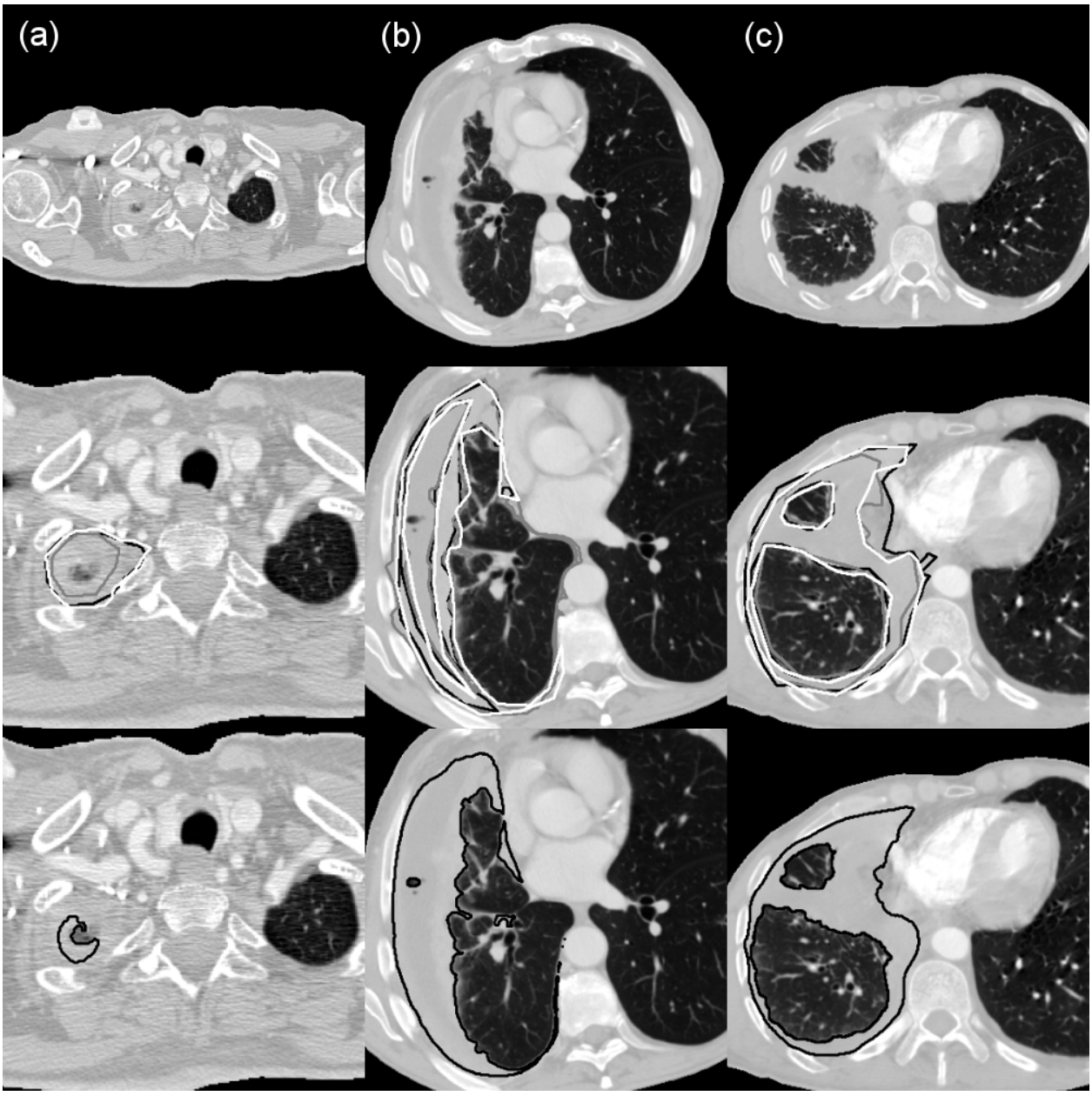

Figure 3.7: Preprocessed CT sections (top), observer reference tumor segmentations (middle; white, gray, and black outlines), and deep CNN tumor segmentations (bottom; black outlines) for three sections from different scans of the two test sets. Sections were selected at random from (a) the bottom 10th percentile (test set 1, average DSC = 0.366), (b) the interquartile range (test set 2, average DSC = 0.647), and (c) the top 10th percentile (test set 1, average DSC = 0.857) of the average DSC value when comparing deep CNN and observer reference segmentations across both test sets. In (b), only three of the five observer segmentations are shown in white, gray and black outlines according to the lowest, highest, and median DSC value for this axial section, respectively.

3.10 Discussion

Scarcity of data is a common issue when implementing machine learning techniques for the medical imaging domain. Studies on deep CNN-based segmentation methods have often applied extensive augmentation to overcome this problem in biomedical applications [143, 144]; furthermore, data augmentation can improve both the performance and generalizability of CNNs to unseen datasets [111, 145]. In the present study, only minimal augmentation was applied to the training set due to the inherent asymmetries of patient anatomy on chest CT scans. As shown in Fig. 3.2, the application of data augmentation to the training sets decreased the overall variance in validation set performance for both sides of the chest; however, Table 3.4 shows that similar optimal deep CNN performance was achieved with and without data augmentation. Deep CNNs trained with data augmentation were ultimately selected for application to the test sets in the present study due to the improved performance and generalizability shown in previous studies on deep CNN-based segmentation.

The present deep CNN-based segmentation method of mesothelioma tumor showed significantly greater overlap with the reference tumor segmentations of all three observers on test set 1 when compared with a previously published segmentation method ("2011 Method"). Furthermore, Bland-Altman plots comparing the segmented tumor area by the deep CNN-based method and the 2011 Method with the average observer-segmented area on test set 1 showed narrower limits of agreement for the deep CNN-based method. These results show an overall superior performance of the present deep CNN-based segmentation method when compared with the 2011 Method and indicate that, in general, deep CNN-based segmentation methods are applicable to the complex task of mesothelioma tumor segmentation on CT scans.

The deep CNN method showed comparable overlap with the five radiologists on test set 2 as with the three observers on test set 1 (see Figs. 3.3 and 3.5); however, Bland-Altman analysis of the relative differences between the deep CNN tumor area and the average observer-segmented tumor area on test set 2 showed increased bias and wider 95% limits of

agreement when compared with test set 1, despite test set 2 including a greater number of axial sections than test set 1 (the width of the estimated 95% limits of agreement is inversely related to the number of samples). This increased bias of the deep CNN method on test set 2 could be partly due to a number of sections in test set 2 exhibiting large pleural effusions combined with the fact that tumor segmentations in the training set of this study did not uniformly exclude pleural effusion of the same laterality as visible tumor. Of the 15 axial CT sections for which the deep CNN-predicted tumor area on test set 2 exceeded 29.4% of the average observer-segmented area (the upper limit of the 95% CI of the mean relative difference between deep CNN-predicted tumor area and average tumor area contoured by the observers), seven sections exhibited large effusions in the pleural space that were both classified as tumor by the deep CNN-based method and excluded from tumor segmentations by all five observers. The median relative difference between deep CNN-predicted tumor area and average observer-segmented area for these seven sections was 124.0% (range: 54.6%–166.3%); leaving these sections out of the analysis reduced the bias between the computerized tumor area and the average observer-segmented tumor area, although the difference in mean bias did not reach statistical significance at the 5% level. This observation suggests that further curation of the training set, possibly combined with additional methods of pixel-wise differentiation between tumor and pleural effusion, will be required in future studies to increase the agreement of deep CNN-predicted tumor area and observer-segmented tumor area.

In this study, deep CNNs were trained separately for the segmentation of mesothelioma tumor in the left and right hemithoraces. Bilateral disease is not common amongst mesothelioma patients [3], and preliminary investigations with deep CNNs trained on a set of scans exhibiting disease in both sides of the chest indicated an increased likelihood of the erroneous classification of pleural thickening in the contralateral hemithorax as tumor. Given the relatively large pool of scans available for training the deep CNNs of this study, it was deemed appropriate to pursue the training of hemithorax-specific CNNs rather than the development

of a post-hoc method for filtering out pixels falsely classified as tumor. While CNNs are designed to be translationally invariant, it was presumed that, given a training set of unilateral tumor segmentations, the CNNs would learn enough global context to avoid erroneously identifying pleural thickening in the contralateral hemithorax as tumor. The specificity of the deep CNNs trained in this study with respect to disease laterality was good; out of the total number of 131 CT sections across the two test sets, only four sections (3%) from the scans of three patients (all with left-hemithorax disease) contained pixels in the contralateral hemithorax erroneously classified as tumor. In two of these sections the erroneous inclusion was due to a large effusion in the right hemithorax (area of contralateral segmented regions: 174 mm^2 and 77 mm^2), in one case it was due to the deep CNN classifying 27 pixels of the outer superior surface of the liver as medial mesothelioma tumor, and in one case the deep CNN included a single pixel of the contralateral pleural space in the tumor segmentation.

The present deep CNN-based segmentation method was completely automated apart from user input on the laterality of disease. The trained deep CNNs can be applied to a new CT scan after minimal preprocessing has taken place: segmentation of the patient's thorax using a simple threshold-based technique and applying the appropriate numerical conversion and linear scaling to the pixel values of the scan. Segmentation of tumor on 100 axial CT sections using the present method took approximately 30 seconds on an Nvidia GeForce GTX Titan GPU (originally released in 2013) with 6 GB of memory.

3.11 Conclusion

In this study, a deep CNN-based method was developed and implemented for the automated segmentation of mesothelioma tumor on CT scans. Deep CNNs were trained separately for the segmentation of disease in the left and right hemithoraces. The present deep CNN-based method showed significantly higher overlap with observer-provided reference segmentations when compared with a previously published method on automated mesothelioma segmentation that utilized a traditional step-wise approach. Future avenues of study for the task of

deep CNN-based mesothelioma segmentation include the training of deep CNNs on larger datasets, the exploration of the application of three-dimensional CNNs to the segmentation task, and the investigation of approaches to distinguish more clearly between tumor pixels and non-tumorous pleural thickening.

This study represents the first exploration of deep CNNs for the segmentation of mesothelioma tumor volume on CT scans; the results from this investigation demonstrated the suitability of deep learning-based methods for this complex task. Pleural effusion is a common presentation on the CT scans of mesothelioma patients; the method presented in this chapter exhibited lower agreement with radiologists on CT sections that exhibited pleural fluid. The improvement and validation of the deep learning-based segmentation of mesothelioma on CT scans that exhibit pleural effusions should be investigated to improve the generalizability and robustness of this class of methods for the segmentation of this disease.

CHAPTER 4
DIFFERENTIATION OF PLEURAL EFFUSION AND MESOTHELIOMA ON CT SCANS USING DEEP CONVOLUTIONAL NEURAL NETWORKS

4.1 Introduction

Data scarcity is a common issue in the training of deep convolutional neural networks (CNNs) for medical imaging-related tasks. As mentioned in Section 1.4.2, researchers have investigated the use of convolutional layers pre-trained on large natural image data sets such as ImageNet [146] for the initialization of deep CNNs applied to clinical tasks through a process called "transfer learning." This strategy can be used to directly extract feature descriptors of medical image regions-of-interest for further classification using a non-deep learning-based machine learning classifier (e.g., in the classification of benign and malignant lesions) or to initialize the layers of a deep CNN that is subsequently fine-tuned for a given clinical task through further training. A majority of mesothelioma patients present with or develop pleural effusion, the presence of which may complicate tumor segmentation on CT scans; therefore, a robust automated segmentation method of mesothelioma tumor should exclude effusion from segmented tumor and the performance of automated methods for the segmentation of mesothelioma tumor should be assessed on a set of scans that exhibit pleural fluid.

In the study presented in Chapter 3, a deep learning-based method was shown to achieve superior performance to a previously published method for the automated segmentation of mesothelioma tumor ("2011 Method") [56] on a distinct test set of CT scans of mesothelioma patients with radiologist-provided tumor contours. This result indicated the suitability of deep CNNs for the segmentation of mesothelioma tumor on CT scans; however, the evaluation of the method from Chapter 3 indicated that improvement in segmentation per-

formance might be achieved through the improved exclusion of pleural effusion from deep CNN-predicted tumor segmentations.

In this study, the deep learning-based segmentation of mesothelioma tumor on CT scans was improved through the initialization of deep CNNs with pre-trained convolutional layers, a revised set of tumor segmentations for training of the networks, and a comprehensive strategy for the evaluation of segmentation performance of the networks during training. These modifications of the deep CNN-based method were intended to aid in the reduction of non-tumorous pixels (in particular, pleural effusion) erroneously classified as tumor by the networks developed in Chapter 3.

The principal aim of this study was to improve the segmentation performance of the deep learning-based method specifically on CT sections that exhibit pleural effusion and tumor through the improved exclusion of pleural fluid from the computerized tumor segmentations. The segmentation performance of the networks trained for this study was assessed on a test set of CT sections of mesothelioma patients who presented with pleural effusion and the combined test sets of the study presented in Chapter 3. The development of a reliable, fully automated method to acquire volumetric segmentations of mesothelioma tumor could provide clinicians with an additional data point to aid them in their patient management-related decisions; furthermore, such a method could eventually lead to an efficient way to estimate tumor bulk for the clinical evaluation of treatment response of this aggressive disease and for the efficient collection of data towards advanced pixel-based analyses of tumor in research studies.

4.2 Scope of Study

Mesothelioma patients commonly present with pleural effusions and atelectatic (i.e., collapsed) lung; these abnormalities have considerable overlap in Hounsfield Unit (HU) values with mesothelioma [18]. Mesothelioma may invade the chest wall, mediastinum and/or the abdomen in late-stage patients [28]. The present study addressed the identification of

mesothelioma tumor in patients who exhibited unilateral disease that had not invaded other structures or organs. Deep CNNs were trained separately for the segmentation of disease in the left and right hemithoraces; results of the present method were compared with (1) manual tumor outlines constructed by three thoracic radiologists on CT sections of two test sets not included in the training set and (2) the output of the method presented in Chapter 3 on the same test sets.

4.3 Data Preprocessing

All CT scans used for training, validation, and testing underwent an in-house thoracic segmentation method developed in Matlab (Mathworks Inc., Natick, MA) to exclude the patient table and surrounding air. All CT sections used for training, validation, and testing were converted from Hounsfield Unit (HU) values to 32-bit floating-point values on the range $[0, 1]$ according to the following piecewise linear scaling:

- Pixels outside the thorax and pixels of value equal to or below -1000 HU were assigned a value of 0, and pixels of value equal to or greater than 240 HU were assigned a value of 1.
- Pixel values on the range (-1000 HU, -700 HU] were linearly scaled to the range $(0, 0.15]$.
- Pixel values on the range (-700 HU, -160 HU] were linearly scaled to the range $(0.15, 0.25]$.
- Pixel values on the range (-160 HU, 240 HU) were linearly scaled to the range $(0.25, 1)$.

This rescaling of pixel values was used to increase the contrast of pixels with values on the "soft-tissue window" range of [-160 HU, 240 HU] [147] relative to the preprocessing settings used for the study presented in Chapter 3 (which employed a single linear scaling of the range [-1000 HU, 400 HU] to the range [0,1]); the use of a soft tissue window has been recommended for the radiologic examination of mesothelioma [47, 148]. The values of the HU ranges and the corresponding floating-point ranges were determined prior to the study through visualization

of a subset of the training set. The time-intensive nature of training deep CNNs precluded the comparison of segmentation performance for multiple preprocessing settings.

Rescaling pixel values on the soft-tissue window range of CT scans of this study to a wider proportion of the floating-point range $[0, 1]$ relative to the study of Chapter 3 was done to increase contrast between tumor and effusion; furthermore, the common use of this windowing setting by radiologists for measurement of mesothelioma indicated that image preprocessing based on this range of pixel values would likely be applicable to a majority of CT scans of mesothelioma patients. Figure 4.1 shows the preprocessing settings of the study presented in Chapter 3 and of the present study applied to a CT section of a mesothelioma patient who presented with pleural effusion and the corresponding histograms of the preprocessed image pixel values. Visual comparison of the different preprocessing settings in Fig. 4.1 shows the increased contrast between tumor and effusion with the piecewise linear scaling employed for the present study.

4.4 Training Set

The training set of the present study consisted of 1) reference contours constructed by an imaging scientist trained in thoracic anatomy (E.G.) on scans of patients who presented with pleural effusion, 2) a revised subset of the training set of the study presented in Chapter 3, 3) scans with reference contours that had been constructed by a radiologist in our lab for a previously published study on the correlation of pathologic and radiologic tumor volume in mesothelioma [58], and 4) sections from two CT scans of the same patient that had reference contours constructed by an oncologist experienced in the treatment of mesothelioma. The training set of the study presented in Chapter 3 used reference tumor contours that were constructed using a semi-automated segmentation method. Tumor segmentations acquired in this manner are somewhat noisy due to the interpolation used in the process to generate segmentations on sections that did not contain observer-provided contours; furthermore, pleural fluid was not consistently excluded from these semi-automatically acquired tumor

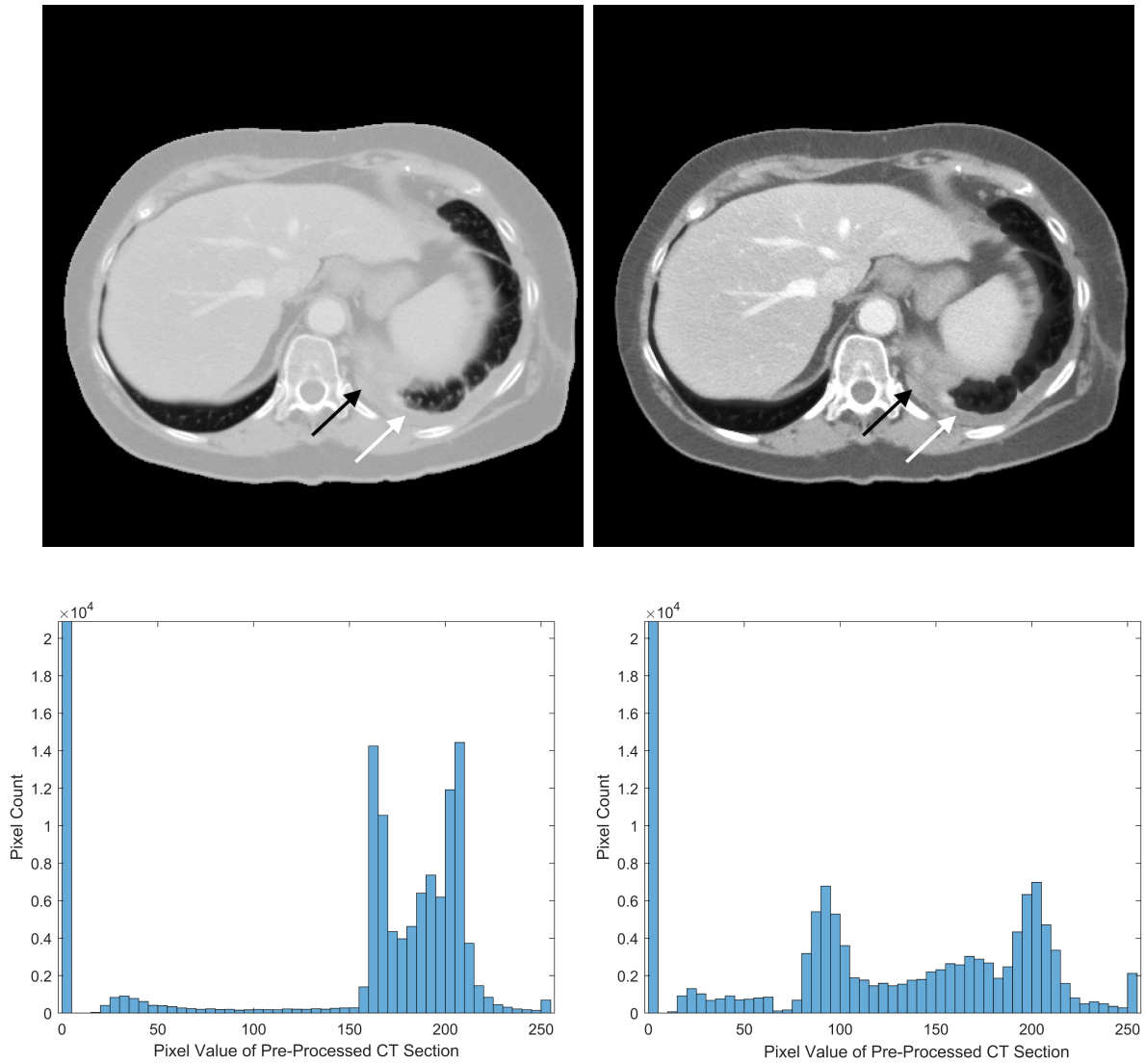

Figure 4.1: Comparison of preprocessing settings on an axial CT section of a mesothelioma patient exhibiting tumor (black arrow) and pleural effusion (white arrow): (top left) CT section with the preprocessing settings from Chapter 3; (top right) CT section with the preprocessing settings of the present study; (bottom left) histogram of image pixel values after the application of the preprocessing settings of Chapter 3; (bottom right) histogram of image pixel values after the application of the preprocessing settings of the present study.

Table 4.1: Characteristics of CT scans used for training the deep CNNs.

Characteristic	Value
Disease laterality	
Left hemithorax	61 patients (76 scans, 2593 sections)
Right hemithorax	65 patients (78 scans, 2486 sections)
Median no. of segmented sections per scan	
Scans with left-sided disease	20.5 (range: 1–126) sections
Scans with right-sided disease	19.5 (range: 1–112) sections
Median slice thickness	
Scans with left-sided disease	3 (range: 0.625–5) mm
Scans with right-sided disease	3 (range: 0.625–5) mm
Median pixel spacing	
Scans with left-sided disease	0.731 (range: 0.561–0.977) mm
Scans with right-sided disease	0.720 (range: 0.588–0.943) mm

segmentations. Preliminary investigations for the present study indicated that inclusion of more than one scan of each patient in the training set might not be beneficial to segmentation performance of the trained networks. These aspects of the training set of the study presented in Chapter 3 thus prompted the revision of the reference tumor contours on that set of scans for the present study.

The final training set consisted of 2593 sections from 73 scans of 61 mesothelioma patients with disease in the left hemithorax and 2486 sections from 78 scans of 65 mesothelioma patients with disease in the right hemithorax (see Table 4.1).

4.5 Test Sets

Two test sets of CT sections of mesothelioma patients with radiologist-provided reference tumor segmentations were used for testing the deep CNNs trained in this study; one of these test sets consisted solely of sections on which both tumor and pleural effusion were present.

The test set specifically created for the present study ("Tumor and Effusion Test Set") consisted of 94 axial CT sections (that all exhibited both tumor and pleural effusion) randomly selected from 46 CT scans of 34 patients not included in the training set of the

Table 4.2: The average interobserver DSC values across all sections for the three Observers of test set 1 of Chapter 3 and the average interobserver DSC value for each observer.

Interobserver DSC	Obs. A	Obs. B	Obs. C	Average DSC
Obs. A	X	0.652	0.814	0.733
Obs. B	0.652	X	0.648	0.650
Obs. C	0.814	0.648	X	0.731

present study. Reference segmentations on this test set were constructed by a radiologist experienced in the measurement of mesothelioma. Sections exhibiting bilateral disease and invasion of anatomic structures adjacent to the pleura were excluded from the test set. To reduce anatomic correlation between sections from the same scan, only a single section randomly selected from within each of the upper-, mid- and lower-thoracic regions of the same scan were included in the test set (for a maximum of three sections per scan); furthermore, all sections from the same scan were separated in the axial direction by at least 1 cm. Of the 94 axial sections of this test set, 40 had left-hemithorax disease and 54 had right-hemithorax disease (see Table 4.4).

The second test set ("Test Set 2") for the present study consisted of 130 CT sections from 43 scans of 43 mesothelioma patients; this test set was created through the combination of the two test sets of the study presented in Chapter 3. A single axial CT section of the test sets of the study in Chapter 3 was excluded from this combined test set due to a chest wall invasion that was overlooked in the study presented in Chapter 3. Table 4.2 lists the average interobserver DSC values for the three observers of test set 1 from Chapter 3; Table 4.3 list the average interobserver DSC values for the five observers of test set 2 from Chapter 3. To simplify the analysis of the present study, a single set of reference tumor contours for Test Set 2 was constructed by using only the reference contours provided by the single observer with the highest average interobserver DSC value on each of the test sets of the study presented in Chapter 3 (Observer A and Observer 4; see Tables 4.2 and 4.3). Of the 130 axial sections of the combined test set, 40 exhibited left-hemithorax disease and 90 exhibited right-hemithorax disease (see Table 4.4).

Table 4.3: The average interobserver DSC values across all sections for the five Observers of test set 2 of Chapter 3 and the average interobserver DSC value for each observer.

Interobserver DSC	Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5	Average DSC
Obs. 1	X	0.720	0.787	0.781	0.783	0.768
Obs. 2	0.720	X	0.720	0.766	0.711	0.729
Obs. 3	0.787	0.720	X	0.787	0.805	0.775
Obs. 4	0.781	0.766	0.787	X	0.766	0.775
Obs. 5	0.783	0.711	0.805	0.766	X	0.766

Table 4.4: Test sets used for segmentation performance assessment in this study.

Characteristic	Value
Tumor and Effusion Test Set	
Sections with left-sided disease	40 sections (43%)
Sections with right-sided disease	54 sections (57%)
Median slice thickness	3 (range: 2.5–4) mm
Median pixel spacing	0.747 (range: 0.557-0.926) mm
Test Set 2	
Sections with left-sided disease	40 sections (31%)
Sections with right-sided disease	90 sections (69%)
Median slice thickness	3 (range: 1–5) mm
Median pixel spacing	0.723 (range: 0.535–0.883) mm

4.6 Deep CNN Architecture

The Visual Geometry Group 16 (VGG16) network architecture, pre-trained on the ImageNet database, was used as the downsampling path of a U-Net deep CNN for the experiments of the present study [116, 149]. The layers of the downsampling path were initialized using weights acquired when the VGG16 was trained with scale-jittering on the ImageNet dataset of natural images (configuration D in the original paper of Simonyan et al. [149]). Initial experiments using the training set of the study presented in Chapter 3 indicated that superior segmentation performance would be achieved through the use of pre-trained convolutional filters when compared with networks trained "from scratch,ÂťÂť as in the study of Chapter 3 [150].

The network architecture is shown in Fig. 4.2. The network accepted as input a 512×512 image matrix and produced a tumor segmentation mask of the same size as the input. Convolutional layers in the network were followed by a rectified linear unit (ReLU) activation function, except at the last layer, for which a sigmoid activation function was used to produce values on the range $(0, 1)$. The continuous output of the network was transformed into binary tumor segmentations using a threshold value of 0.5, determined prior to the start of the study. A 2×2 max pooling operation with stride 2 was used to implement the downsampling of the feature matrix at each level of the downsampling path. The number of feature channels was doubled at each downsampling step, starting with 64 channels at the input level of the network. A dropout layer of probability 0.5 was used during latter stages of the downsampling path to prevent overfitting. At each level of the upsampling path, a two-dimensional upsampling operation using nearest-neighbor interpolation was applied to the feature matrix, and the resulting feature map was concatenated with the feature map from the corresponding level of the downsampling path.

Network loss during training was calculated as the binary cross-entropy L averaged over all pixels of each deep CNN-predicted segmentation and the corresponding reference tumor

segmentation:

$$L(t_i, p_i) = -\left(t_i \log\left(p_i\right) + \left(1 - t_i\right) \log\left(1 - p_i\right)\right) \ , \tag{4.1}$$

where t_i is an indicator variable taking the value 1 if the reference classification of pixel i is tumor and 0 otherwise, and p_i is the deep CNN-predicted probability that pixel i is tumor ($p_i = 1$) or background ($p_i = 0$). The Adam method was used to optimize the network during training using a learning rate of 10^{-5}, chosen after initial investigations on a subset of the training set. The deep CNN architecture was implemented with the Keras and Tensorflow deep learning frameworks. Experiments were run using a batch size of 1 on a scientific computing cluster at the University of Chicago using Nvidia GeForce GTX Titan and Nvidia Tesla K20c Kepler-class graphics processing units (GPUs; Nvidia, Santa Clara, CA).

4.7 Experiments

Deep CNNs were trained separately on sections and reference segmentations of mesothelioma patients with visible disease in the left and right hemithorax. Validation sets were used to select the optimal deep CNN of each hemithorax for application to the test set. For the left hemithorax, 10 scans from 9 patients were selected from the training set and used as a validation set during training; for the right hemithorax, 10 scans from 10 patients were selected from the training set and used as a validation set during training.

To investigate the effect of the relative proportion of the training set that exhibited both tumor and effusion and that exhibited tumor with no apparent effusion on the segmentation performance of the optimal networks, the training and validation set of each hemithorax was divided into subsets of 1) sections that exhibited tumor with no apparent effusion ("tumor only") and 2) sections that exhibited both tumor and effusion ("tumor and effusion"). Table 4.5 shows the number of sections in each subset of the training and validation set of each hemithorax; in the training set of each side of the chest, there were approximately four

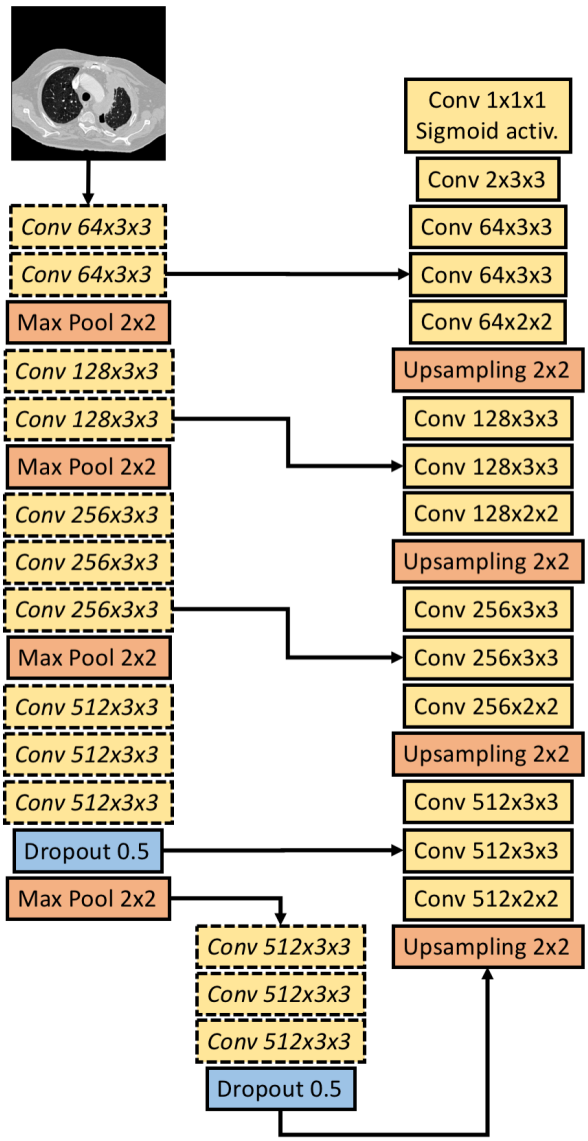

Figure 4.2: The VGG16/U-Net deep CNN architecture of the present study. Dashed lines indicate VGG16 layers pre-trained on the ImageNet database. Arrows indicate the flow of the input matrix and feature matrix through the network. Convolutional layers are labelled as "Conv" and a triplet of numbers that represents the number of feature channels, height of the convolution window, and width of the convolution window, respectively. A dropout probability of 0.5 was used during training. ReLU activation functions were used following convolutional layers, except where noted. Upsampling was implemented through nearest neighbor interpolation.

times as many sections that exhibited tumor without apparent effusion than sections that exhibited both tumor and effusion. To determine the optimal relative proportion of the two classes of sections (i.e., "tumor only" and "tumor and effusion") in the training set of each hemithorax, the values of 1:1, 2:1 and 4:1 were explored for the relative proportion of the two classes during training; the validation set of each hemithorax was used to determine the optimal relative frequency of the two classes before a single deep CNN in each hemithorax was chosen for application to the test sets.

Three metrics of segmentation performance were calculated on the validation set of each hemithorax during training: the median DSC, the median average Hausdorff distance (AHD; also known as the modified Hausdorff distance) [151], and the ratio P of the total number of predicted tumor pixels to the total number of reference tumor pixels. The DSC was calculated as the ratio of twice the area of intersection of the segmentations under comparison to the sum of the area of the segmentations [137, 138]:

$$\mathrm{DSC} = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}\,, \tag{4.2}$$

where $|S_1|$ and $|S_2|$ represent the respective area of each segmentation and $|S_1 \cap S_2|$ represents the area of the intersection of the two segmentations. AHD is an evaluation metric for a pair of segmentations that takes into account the spatial location of the segmentation boundaries [138, 151]:

$$\mathrm{AHD}(A, B) = \max\left(\overline{d}(A, B), \overline{d}(B, A)\right) \quad , \tag{4.3}$$

where A and B represent the two tumor contours (i.e., tumor segmentation outlines) under comparison and $\overline{d}(A, B)$ is defined as the average Euclidean distance to the nearest point on contour B from each point on contour A. The AHD metric has been found to be more robust in the presence of outliers relative to the original Hausdorff distance metric [151]. These three metrics were calculated on the validation set approximately every 1000 network

updates during training. Deep CNNs were trained for 3×10^5 updates or until performance metric values on the validation set indicated that the deep CNN had started to overfit the training set.

Minimal data augmentation was applied to the training set due to the inherent asymmetry of chest anatomy. Data augmentation for the study presented in Chapter 3 was applied to the training set of that study through a random scaling in the range $[0.95, 1.05]$ and a random rotation in the range $[-5°, +5°]$; these random values of rotation and scaling were determined separately for each training sample and applied to the training set before each epoch of training. This augmentation approach provided the networks of Chapter 3 with a uniquely augmented training set for each iteration over the data; however, it was found that this randomized approach could produce CT sections of unrealistic appearance. For the present study, a rotation of either $-10°$ or $+10°$ and a scaling of either 0.9 or 1.1 were each manually selected for each CT scan of the training set to ensure a realistic appearance of the augmented CT sections; all combinations of the selected rotation angle and scaling factor were applied to each CT section of the respective scan. The number of sections in each training set of the present study was thus artificially increased by a factor of four. The values of rotation and scaling were determined by visualizing different rotation angles and scaling values on example CT sections from the training set.

4.8 Selection of CNNs for Application to the Test Sets

Median AHD was found to be a more representative metric of overall segmentation performance during training than median DSC. Figure 4.3 shows a scatter plot of the median DSC and median AHD values obtained for the "tumor only" and "tumor and effusion" subsets of the validation set of the left and right hemithoraces, for the case of 4:1 and 1:1 relative frequency of "tumor only" to "tumor and effusion" CT sections in the training set for the left and right hemithorax, respectively. The general trend shown in Fig. 4.3, for which lower median AHD values were associated with relatively high median DSC values and higher median DSC

Table 4.5: Division of training and validation sets of each hemithorax into sections that exhibit tumor without apparent effusion ("tumor only") and sections that exhibit tumor with apparent effusion ("tumor and effusion"). No scans of the same patient formed a part of both the training and validation set of a given hemithorax.

Characteristic	No. of Scans	Median No. of Sections per Scan
Left hemithorax, training set		
Tumor only	33 (2138 sections)	58 (range: 8–126)
Tumor and effusion	54 (525 sections)	7 (range:1–68)
Left hemithorax, validation set		
Tumor only	4 (275 sections)	70 (range: 58–77)
Tumor and effusion	8 (97 sections)	9 (range: 3–33)
Right hemithorax, training set		
Tumor only	35 (2047 sections)	59 (range: 14–112)
Tumor and effusion	50 (520 sections)	6 (range: 1–48)
Right hemithorax, validation set		
Tumor only	4 (215 sections)	56 (range: 16–88)
Tumor and effusion	8 (101 sections)	6.5 (range: 1–38)

values were not necessarily associated with low median AHD values, was evident on plots of median DSC as a function of median AHD for the validation sets of both hemithoraces. The minimum median AHD on the validation set of each hemithorax was therefore selected as the segmentation performance metric for the selection of CNNs for application to the test sets.

Counterintuitively, it was found that an overall better segmentation performance was achieved across both subsets (i.e., "tumor only" and "tumor and effusion") of the validation sets when the subset for which the minimum median AHD achieved was higher (i.e., the subset on which worse segmentation performance was achieved) was used to select the optimal network for application to the test sets (as opposed to, for example, selecting the network that corresponded to the minimum average median AHD across both subsets of each validation set).

4.9 Interobserver Agreement for CT Sections that Exhibit Tumor and Effusion

To evaluate interobserver agreement in the manual segmentation of mesothelioma tumor on CT sections that exhibit pleural effusion, the mean interobserver DSC value between five radiologists (i.e., the mean of the ten interobserver comparisons made for each section) on a set of 69 axial CT sections from 27 scans of 27 patients was compared between (1) sections for which at least one observer excluded an area of pleural effusion from the tumor contours and (2) sections for which none of the five observers excluded an area of pleural effusion from tumor contours. These images formed a part of Test Set 2 of the present study and were used in a previously published study on observer variability in mesothelioma tumor area measurements [59]. This assessment of interobserver agreement was made only across a subset of the images of Test Set 2 due to the two different sets of Observers who segmented tumor on the original test sets that made up Test Set 2 (see Section 4.5).

4.10 Statistical Analysis

The two-sided Wilcoxon signed-rank test was used to test the null hypothesis that the distributions of DSC values and AHD values were identical for the present method and the method from Chapter 3 when compared with reference tumor segmentations of the two test sets of this study. The two-sided Wilcoxon rank-sum test was used to test the null hypothesis that the distributions of average interobserver DSC values were identical for 1) sections on which at least one out of five radiologists excluded an area of effusion from mesothelioma tumor contours and 2) sections on which none of five radiologists excluded an area of effusion from tumor contours on a set of 69 CT sections that formed part of a previously published study on interobserver agreement in measurements of mesothelioma tumor area [59]. The Bonferroni-Holm correction was applied to the significance level of all statistical tests to account for the number of statistical comparisons made; since five statistical comparisons were

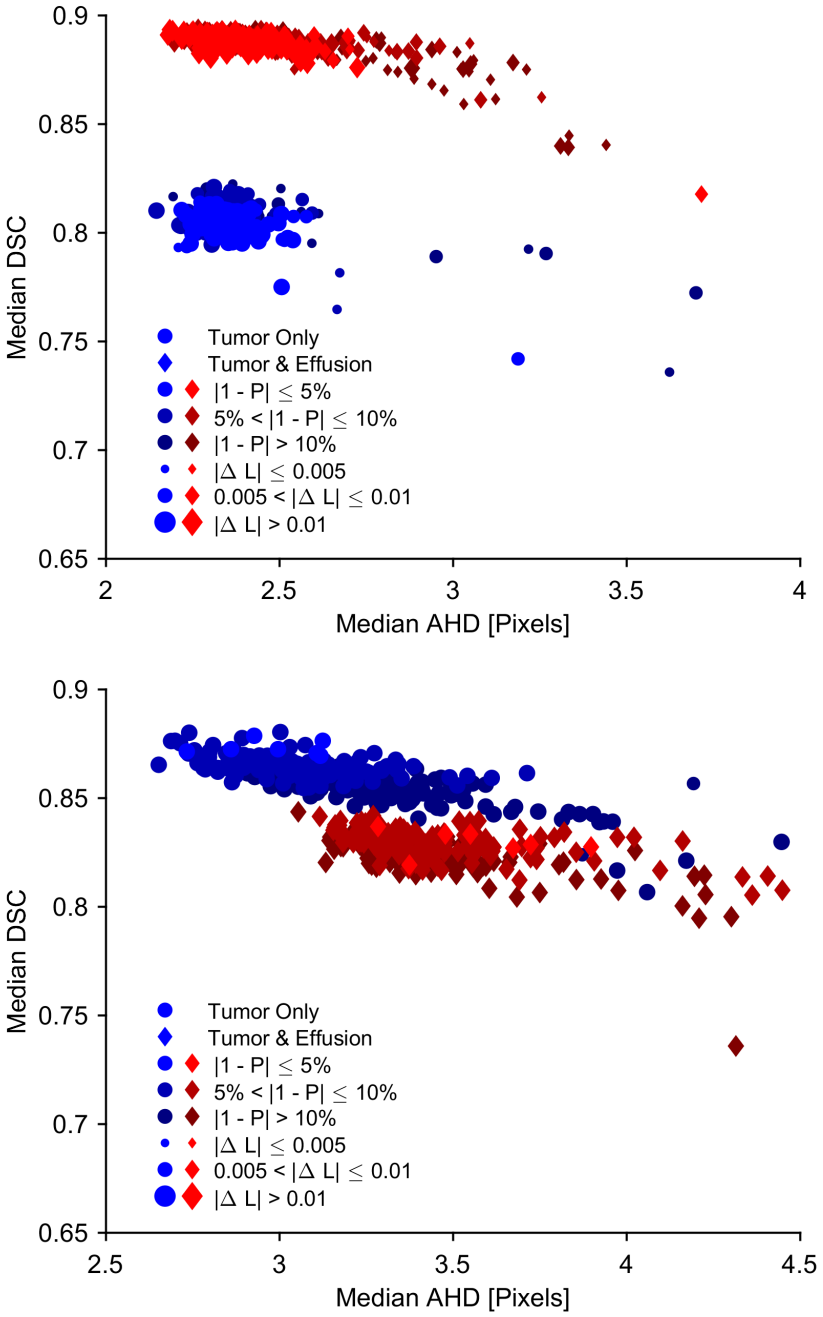

Figure 4.3: Median DSC as a function of the median AHD obtained on the "tumor only" and "tumor and effusion" subsets of the validation set of each hemithorax: (top) the left hemithorax, for a 4:1 relative frequency of tumor only and tumor and effusion sections in the training set, (bottom) the right hemithorax, for a 1:1 relative frequency of tumor only and tumor and effusion sections in the training set. The color shade of the markers represents the absolute deviation from 1 of the ratio P of deep CNN tumor pixels and reference tumor pixels, taken as an average across each subset of the respective validation set. The size of the markers represents the absolute value of the difference ΔL between the average binary cross-entropy loss L across the whole validation set and the average loss L across the training set of each hemithorax at each update of the network.

made, the nominal significance level was adjusted to $\alpha' = 0.05/5 = 0.01$ for the statistical test that achieved the lowest p value of this study, with subsequently higher nominal significance levels $\alpha' = \frac{\alpha}{m-1}, \frac{\alpha}{m-2}, \ldots$ for other statistical tests in order of increasing p-value [132]. Statistical comparisons were made using Matlab.

Bland-Altman plots were used to evaluate agreement between (1) the tumor area segmented by the present method and observer-segmented tumor area on both test sets and (2) between the tumor area segmented by the method presented in Chapter 3 and the tumor area segmented by the observers on the two test sets. Absolute differences in the segmented area of the computerized methods and observer-segmented area were found to correlate with the average segmented tumor area of the segmentation approaches being compared, violating the normality assumption for calculation of 95% limits of agreement according to the Bland-Altman method. Therefore, the 95% limits of agreement were estimated using relative differences in segmented area as $d \pm 1.96s$, where d was the mean and s was the standard deviation of the relative differences between the two segmentation approaches being compared [133]. The standard error (SE) of d was estimated as $\sqrt{s^2/n}$, and the SE of the 95% limits of agreement was estimated as $\sqrt{3s^2/n}$, where n was the number of samples. 95% confidence intervals (CIs) for d and the 95% limits of agreement were found by adding and subtracting 1.96×SE from each value in question [134].

4.11 Results

For the left hemithorax, the minimum median AHD achieved was higher on the "tumor and effusion" validation set; a 4:1 relative frequency of "tumor only" and "tumor and effusion" subsets of the training set was found to achieve the lowest median AHD on this subset of the validation set. For the right hemithorax, the minimum median AHD achieved was higher on the "tumor only" validation set; a 1:1 relative frequency of "tumor only" and "tumor and effusion" subsets of the training set was found to achieve the lowest median AHD on this subset of the validation set. All results presented in this section refer to these two cases for

the respective side of the chest.

4.11.1 Training

The average binary cross-entropy loss L on the training and validation set of each hemithorax and the median DSC and median AHD across the validation set of each hemithorax are shown in Fig. 4.4. The solid lines in Fig. 4.4 indicate the average loss on the training set and the average loss, average median DSC, and average median AHD across the "tumor only" and "tumor and effusion" validation sets for each hemithorax. The shaded areas in Fig. 4.4 indicate the range of the loss, median DSC, and median AHD across the "tumor only" and "tumor and effusion" validation sets of each hemithorax. Table 4.6 lists the minimum median AHD achieved on the validation subset of each hemithorax that was found to achieve a higher minimum median AHD and the corresponding median DSC, loss, and pixel ratio for both validation sets at the corresponding training update.

4.11.2 Tumor and Effusion Test Set

Figure 4.5 shows boxplots of DSC values and AHD values obtained when comparing the predicted tumor segmentations of the present deep CNN method and the method from Chapter 3 with the reference segmentations on the Tumor and Effusion Test Set. The median DSC and median AHD for the present method on this test set were 0.690 (range: 0.070–0.936) and 5.1 mm (range: 0.9–59.1 mm), respectively. The median DSC and median AHD for the Chapter 3 method on the same CT sections were 0.499 (range: 0.055–0.907) and 6.3 mm (range: 2.0–57.0 mm), respectively. Differences in the distributions of DSC values ($p < 0.00001$) and AHD values ($p < 0.001$) between the two methods on this test set were found to be statistically significant using the two-sided Wilcoxon signed-rank test.

Figure 4.6(a) shows a Bland-Altman plot of the relative differences in segmented tumor area by the present deep CNN method and the observer-segmented tumor area on the Tumor and Effusion Test Set. The mean relative difference in segmented tumor area between the

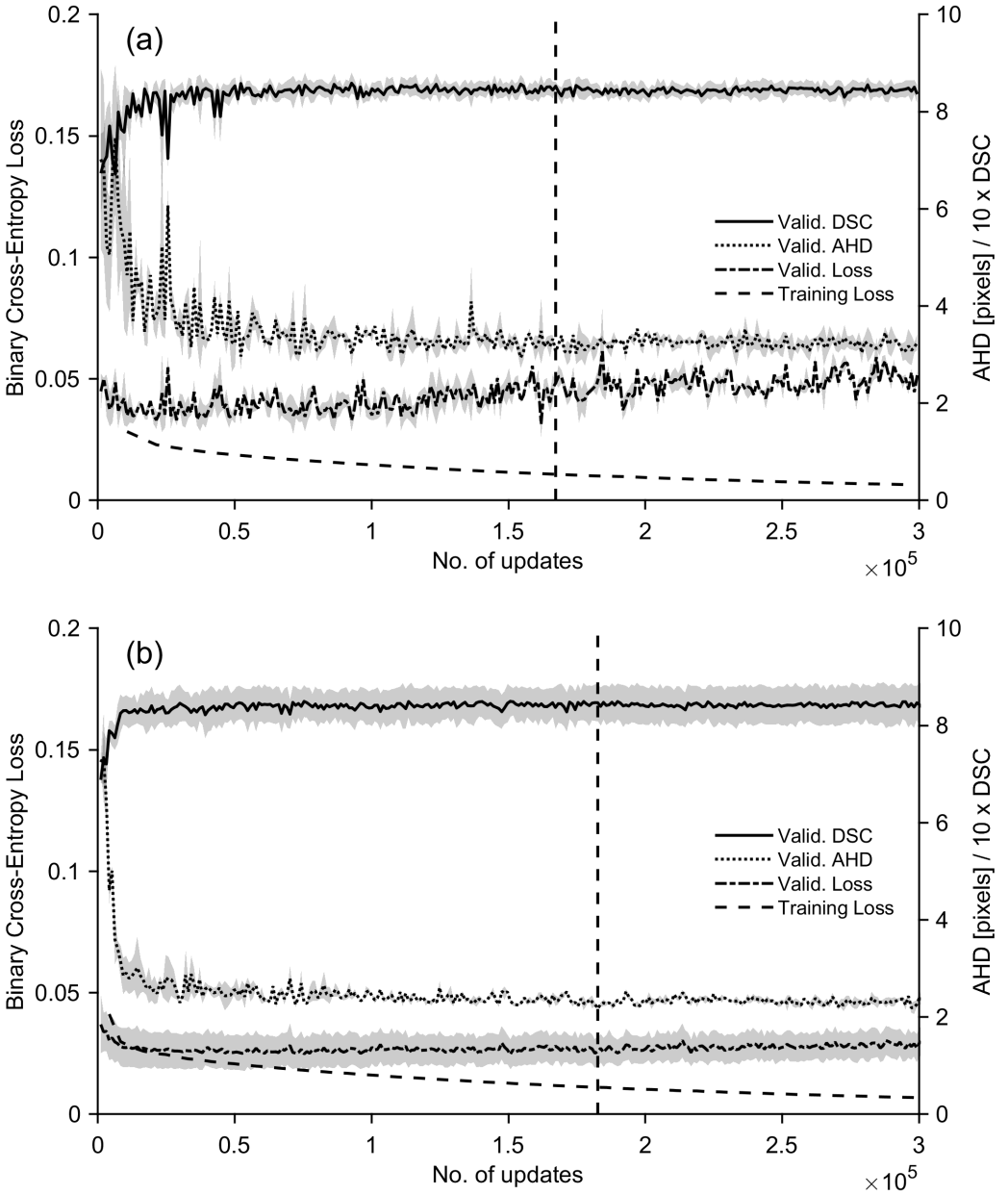

Figure 4.4: Average loss L on the training set and average loss L, median DSC, and median AHD on the validation sets during training (a) of the left-hemithorax deep CNN and (b) of the right-hemithorax deep CNN. Solid lines indicate average values on the training set and across the "tumor only" and "tumor and effusion" validation sets of each hemithorax. Shaded areas indicate the range of the average loss, median DSC, and median AHD across the "tumor only" and "tumor and effusion" validation sets of each hemithorax. Median DSC values are shown with a scaling factor of 10 for visual clarity. Validation set performance was assessed approximately every 1000 updates. The vertical dashed lines indicate the training updates after which the deep CNNs were applied to the test sets.

Table 4.6: Minimum median AHD value achieved on the "tumor and effusion" validation set of the left hemithorax (the lower performing of the two subsets of the validation set) and the minimum median AHD value achieved on the "tumor only" validation set of the right hemithorax (the lower performing of the two subsets of the validation set) and the corresponding average loss L, median DSC value, median AHD value, and the ratio P of the number of predicted tumor pixels and reference tumor pixels for both validation sets of each hemithorax during training.

Hemithorax	Validation Set	Metric	Training Update	Value
Left	Tumor and Effusion	Minimum Median AHD	167,236	3.05 pixels
		Average L	-	0.048
		Median DSC	-	0.844
		P	-	0.95
	Tumor Only	Median AHD	-	3.07 pixels
		Average L	-	0.048
		Median DSC	-	0.844
		P	-	0.82
Right	Tumor Only	Minimum Median AHD	182,620	2.12 pixels
		Average L	-	0.02
		Median DSC	-	0.809
		P	-	1
	Tumor and Effusion	Median AHD	-	2.22 pixels
		Average L	-	0.032
		Median DSC	-	0.885
		P	-	1.08

present method and the observer-segmented area was -8.2% (95% CI: -17.5% to 1.1%) with 95% limits of agreement [-96.3%, 79.9%] (95% CIs: -112.1% to -80.6% and 64.2% to 95.6% for the lower and upper limit, respectively). Figure 4.6(b) shows a Bland-Altman plot of the relative differences in segmented tumor area by the Chapter 3 method and the tumor area segmented by the Observer on the Tumor and Effusion Test Set. The mean relative difference in segmented tumor area between the Chapter 3 method and the observer-segmented area was 68.6% (95% CI: 58.4% to 78.7%) with 95% limits of agreement [-27.7%, 164.9%] (95% CIs: -44.9% to -10.5% and 147.7% to 182.1% for the lower and upper limit, respectively).

Figure 4.7 shows the preprocessed CT sections, observer reference segmentations, and predicted tumor segmentations of the present method for three example CT sections selected at random from the lowest 10th percentile, the interquartile range, and the top 10th percentile of the DSC values found when comparing predicted tumor segmentations of the present method with reference segmentations on the Tumor and Effusion Test Set.

4.11.3 Test Set 2

Figure 4.10 shows boxplots of DSC values and AHD values obtained when comparing the predicted tumor segmentations of the present method and the method from Chapter 3 with the set of reference tumor segmentations on Test Set 2. The median DSC value and median AHD value for the present method on Test Set 2 was 0.780 (range: 0.175–0.927) and 2.9 mm (range: 0.7–50.7 mm), respectively. The median DSC value and median AHD value for the Chapter 3 method on Test Set 2 was 0.764 (range: 0.108–0.938) and 3.3 mm (range: 0.8–56.9 mm), respectively. The difference in the distributions of AHD values ($p = 0.008$) between the two methods on this test set was found to be statistically significant using the two-sided Wilcoxon signed-rank test; the difference in DSC value distributions did not reach statistical significance on this test set using the two-sided Wilcoxon signed-rank test ($p = 0.23$).

Figure 4.9(a) shows a Bland-Altman plot of the relative differences in segmented tumor area by the present method and the average observer-segmented tumor area on Test Set

2. The mean relative difference in segmented tumor area by the present method and the observer-segmented area was -17.4% (95% CI: -23.1% to -11.7%) with 95% limits of agreement [-80.8%, 46.0%] (95% CIs: -90.5% to -71.2% and 36.4% to 55.6% for the lower and upper limit, respectively). Figure 4.9(b) shows a Bland-Altman plot of the relative differences in segmented tumor area by the Chapter 3 method and the observer-segmented tumor area on Test Set 2. The mean relative difference in segmented tumor area between the Chapter 3 method and the observer-segmented area on this test set was 11.8% (95% CI: 3.3% to 20.2%) with 95% limits of agreement [-83.1%, 106.6%] (95% CIs: -97.5% to -68.7% and 92.2% to 121.0% for the lower and upper limit, respectively).

4.11.4 Interobserver Agreement for CT Sections that Exhibit Tumor and Effusion

Out of the 69 CT sections in the set of images used in the assessment of interobserver agreement, one or more of the five Observers excluded pleural effusion from tumor contours on 26 sections (38%); the mean of the average interobserver DSC values on these 26 sections was 0.712 (median 0.743, range 0.517–0.853). The mean of the average interobserver DSC value on the 43 sections of this set for which none of the five Observers excluded pleural effusion from tumor contours was 0.757 (median 0.779, range 0.517–0.915). The difference in the DSC distributions for these two subsets did not reach statistical significance using the two-sided Wilcoxon rank-sum test ($p = 0.07$).

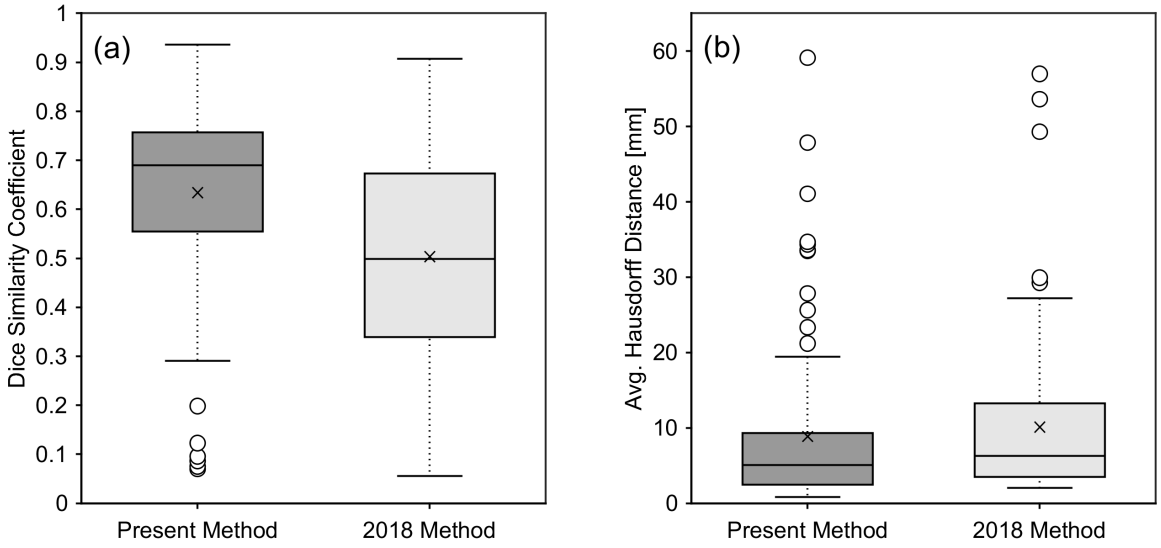

Figure 4.5: Boxplots showing (a) DSC values and (b) AHD values obtained when comparing predicted tumor segmentations by the present deep CNN-based method and the method presented in Chapter 3 with radiologist-acquired reference tumor segmentations of the Tumor and Effusion Test Set of this study. Horizontal lines inside boxes indicate the median value of each distribution; crosses indicate the mean value of each distribution.

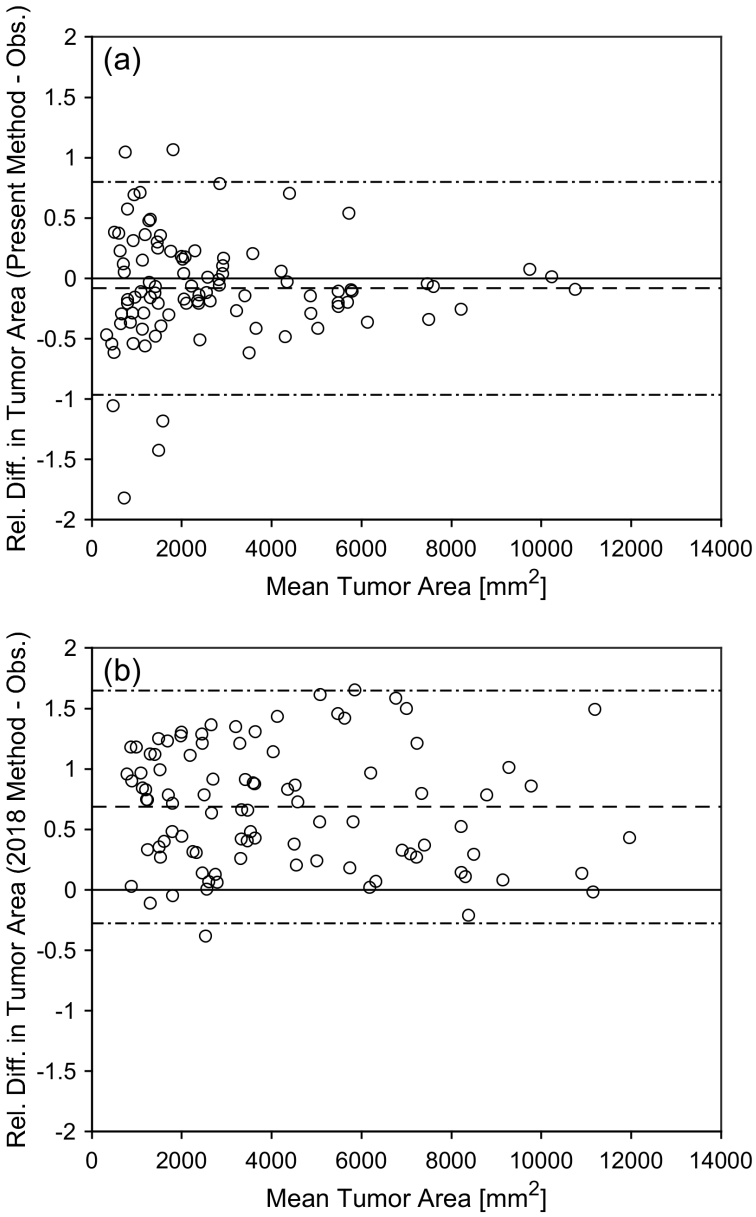

Figure 4.6: Bland-Altman plots showing (a) the relative differences between the segmented tumor area of the present method and the observer-segmented tumor area on the Tumor and Effusion Test Set, and (b) the relative differences between the segmented tumor area of the method presented in Chapter 3 and the observer-segmented tumor area on the Tumor and Effusion Test Set. Means of relative differences and 95% limits of agreement shown as dashed lines.

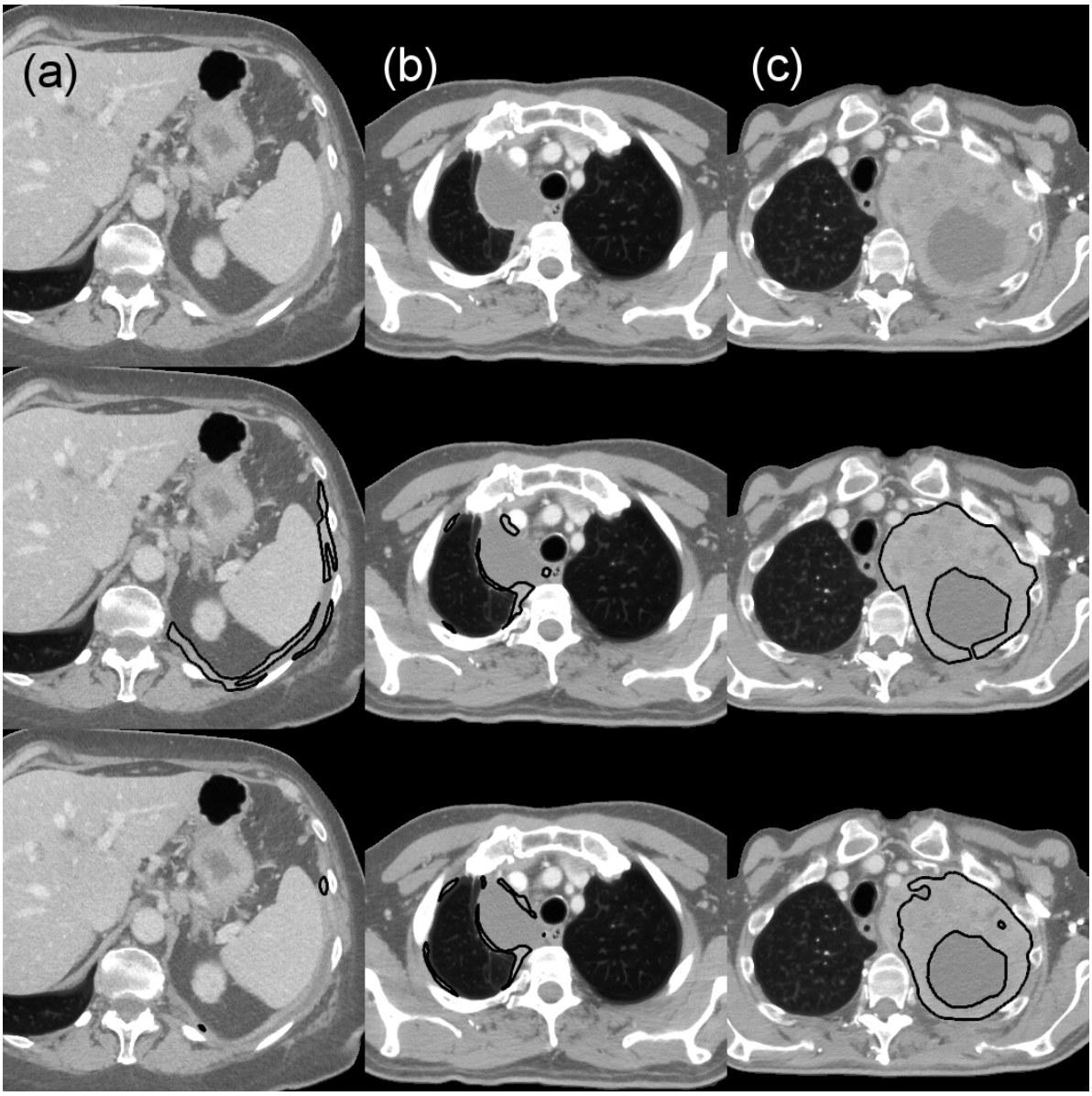

Figure 4.7: Preprocessed CT sections (top), observer reference tumor segmentations (middle; black outlines), and predicted tumor segmentations by the present method (bottom; black outlines) for three CT sections from different scans of the Tumor and Effusion Test Set. Sections were selected at random from (a) the bottom 10th percentile (DSC = 0.086, AHD = 33.6 mm), (b) the interquartile range (DSC = 0.619, AHD = 4.5 mm), and (c) the top 10th percentile (DSC = 0.880, AHD = 3.0 mm) of the DSC values obtained when comparing predicted tumor segmentations of the present method and observer reference segmentations.

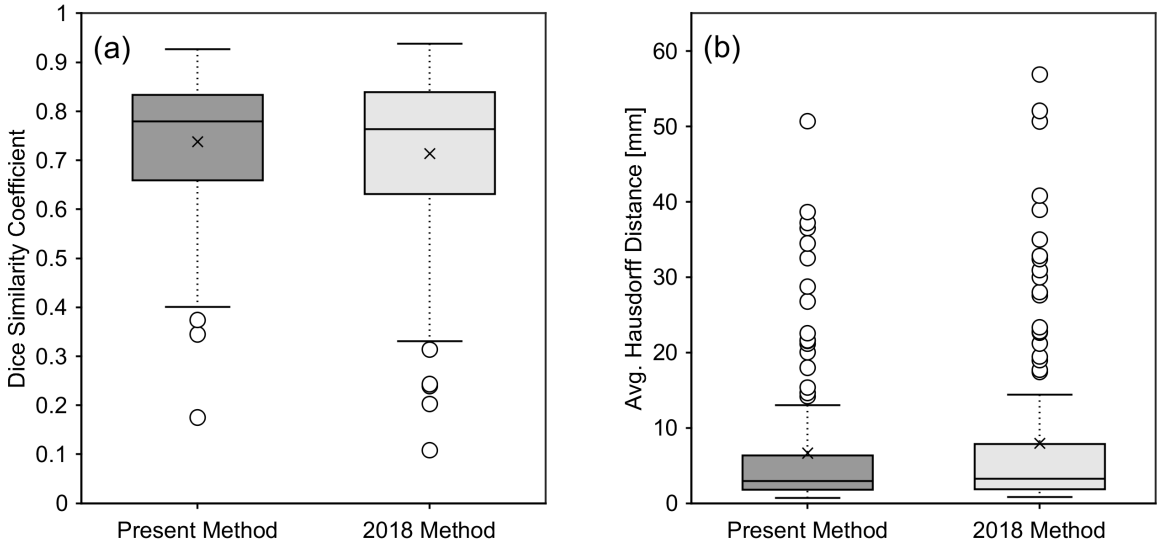

Figure 4.8: Boxplots showing (a) DSC values and (b) AHD values obtained when comparing predicted tumor segmentations by the present deep CNN-based method and the method presented in Chapter 3 with radiologist-acquired reference tumor segmentations on Test Set 2 of this study. Horizontal lines inside boxes indicate the median value of each distribution; crosses indicate the mean value of each distribution.

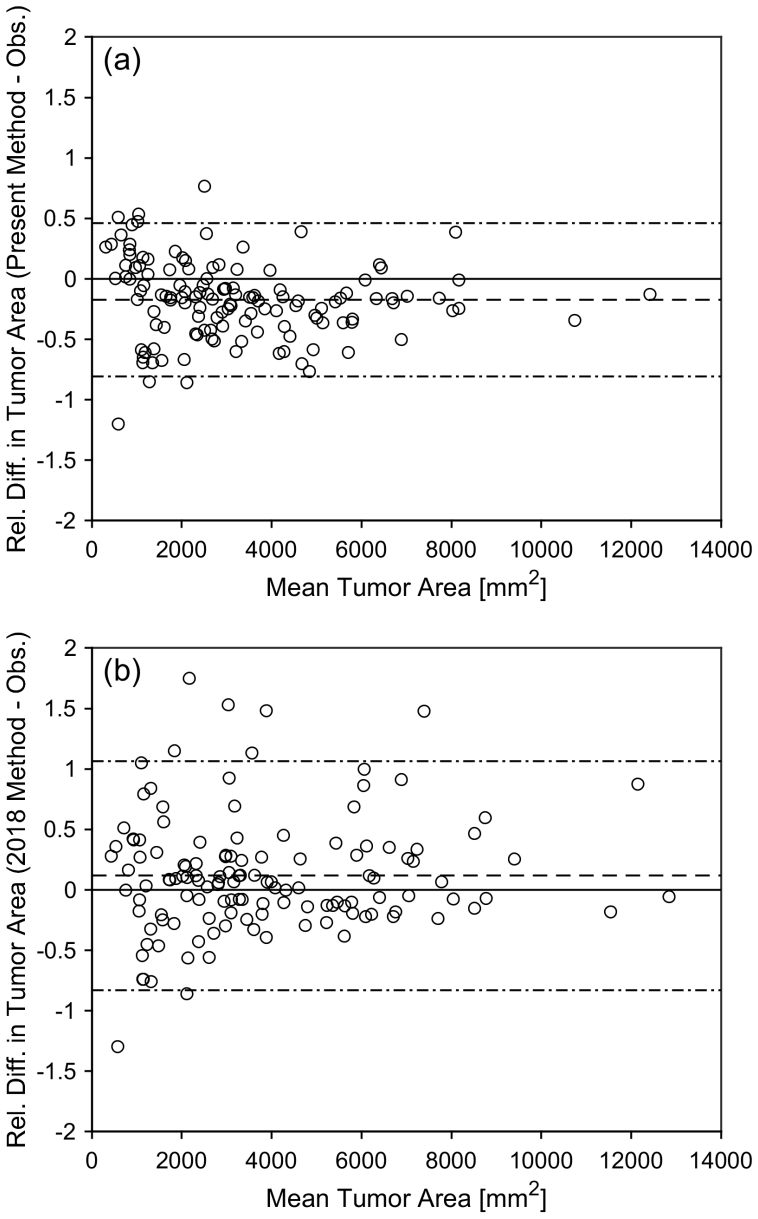

Figure 4.9: Bland-Altman plots showing (a) the relative differences between the segmented tumor area of the present method and the observer-segmented tumor area on Test Set 2 and (b) the relative differences between the segmented tumor area of the method presented in Chapter 3 and the observer-segmented tumor area on Test Set 2. Means of relative differences and 95% limits of agreement shown as dashed lines.

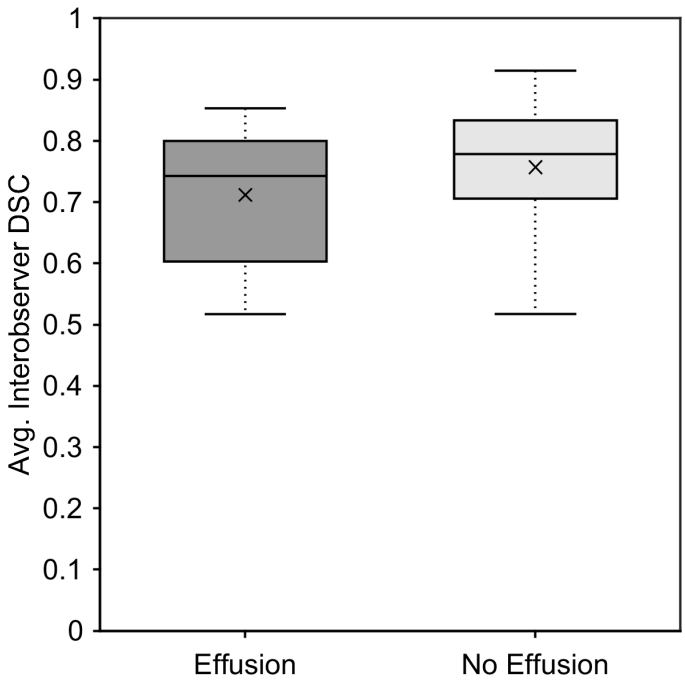

Figure 4.10: Boxplot showing the average interobserver DSC values across five radiologists on CT sections for which at least one radiologist excluded an area of effusion from tumor contours ("Effusion", $n = 26$) and for which no radiologist excluded an area of effusion from tumor contours ("No Effusion", $n = 43$) on a set of 69 CT sections (from 27 scans of 27 patients) that formed a part of Test Set 2 of the present study and of a previously published study on interobserver variability in measurement of mesothelioma tumor area on CT scans. Horizontal lines inside boxes indicate the median value of each distribution; crosses indicate the mean value of each distribution.

4.12 Discussion

Mesothelioma patients commonly present with pleural effusion on imaging; a majority of patients with this disease present with effusion at initial diagnosis [17, 152]. The volumetric segmentation of mesothelioma tumor requires the exclusion of fluid from the segmented tumor volume. The deep CNN-based segmentation method presented in Chapter 3 showed a significantly improved segmentation performance when compared with a previous step-wise segmentation method for mesothelioma [56]; however, the study in Chapter 3 did not adequately exclude pleural effusion from tumor contours. The deep CNNs trained in the present study showed significantly greater overlap with radiologist-provided reference tumor segmentations on a test set of 94 CT sections ("Tumor and Effusion" test set) that all exhibited both tumor and pleural effusion when compared with the study presented in Chapter 3. The agreement of computerized tumor contours and observer-provided reference tumor contours on this test set, as evaluated using the AHD metric, was found to be significantly higher for the present method when compared with the Chapter 3 method. Bland-Altman plots comparing the segmented tumor area by the present method and the Chapter 3 method with observer-segmented tumor area on the Tumor and Effusion test set showed a reduction in bias for the present method when compared with the Chapter 3 method and the 95% CI for the mean relative bias of the present method included 0. These results show a significant improvement in the performance of the present segmentation method when compared with the study presented in Chapter 3 for the task of segmenting mesothelioma tumor on CT scans that exhibit pleural fluid.

The presence of pleural effusion on the CT scans of mesothelioma patients may increase observer variability in the task of mesothelioma segmentation due to the potentially unclear boundaries of tumor and fluid and the overlap in HU values between tumor and pleural fluid. A limitation of the present study was the inability to obtain interobserver comparisons on the "Tumor and Effusion" test set of this study. To estimate the effect of pleural effusion on tumor contour overlap between radiologists in the task of mesothelioma segmentation on

CT scans, the mean interobserver DSC value between five radiologists on a set of 69 axial CT sections was compared between (1) sections for which at least one radiologist excluded an area of pleural effusion from tumor contours and (2) sections for which none of the five radiologists excluded an area of pleural effusion from tumor contours. These images formed a part of Test Set 2 of the present study and were used in a previously published study on observer variability in mesothelioma tumor area measurements [59]. The difference in the DSC distributions for these two subsets did not reach statistical significance ($p = 0.07$); however, the lower mean interobserver DSC value for sections on which pleural effusion appeared on this set of CT sections suggests that the concurrent presence of pleural fluid and mesothelioma tumor results in lower radiologist agreement for the task of mesothelioma tumor segmentation on CT scans.

The second test set of this study ("Test Set 2") included sections that were used to evaluate the segmentation performance of the method presented in Chapter 3. The present method did not show a significantly higher overlap with the set of radiologist-provided reference tumor segmentations on this test set when compared with the Chapter 3 method; however, the present method did achieve a significantly lower median AHD when compared with the method from Chapter 3. Bland-Altman analysis of the predicted tumor area by the two computerized methods on Test Set 2 showed a slight negative mean bias in predicted tumor area for the present method; however, the 95% limits of agreement for the relative difference in computerized tumor area and observer-segmented tumor area were narrower for the present method when compared with the Chapter 3 method. Across all observers who provided reference tumor segmentations on this test set in the previously published study, there were 39 sections (30%) for which at least one observer excluded an area of apparent effusion from the tumor segmentation. The superior segmentation performance of the present method on this set of images shows that, despite the method's principal aim being the improvement of mesothelioma tumor segmentation on scans that exhibit pleural effusion, the performance of this method remains adequate across a test set for which the

majority of cases do not exhibit pleural effusion.

The present study trained deep CNNs separately for the segmentation of disease in the left and right hemithorax. Across the 94 axial CT sections of the Tumor and Effusion test set of this study, there were no pixels erroneously predicted as tumor in the contralateral hemithorax. For the 130 sections of Test Set 2, there were three sections on which pixels of the contralateral hemithorax were erroneously included in the predicted segmentation by the present method. In one of these cases, 211 pixels (corresponding to a volume of 32 mm^3) of the descending aorta and hilar vessels were identified as tumor on a non-contrast-enhanced scan; in the other two cases, parts of a contralateral pleural effusion on two sections of the same CT scan were erroneously identified as tumor (71 and 833 pixels, corresponding to a volume of 24 mm^3 and 282 mm^3, respectively).

The method presented in this study did not incorporate three-dimensional convolutional filters for the automated segmentation of mesothelioma tumor; the improvement in segmentation performance when compared with the method presented in Chapter 3 was achieved through a revision of the training set of the previous study, an extensive validation methodology, and a more varied set of training sections that exhibited tumor with pleural effusion. Results on the validation sets of this study indicate that axial context could aid in the deep learning-based segmentation of mesothelioma on sections where tumor is located adjacent to soft-tissue structures (e.g., medial tumor). The training of three-dimensional CNNs requires higher-memory GPUs due to the larger image volumes that are processed during training and the increase in the number of parameters associated with three-dimensional convolutional filters. Other deep CNN-based segmentation studies have employed image subsampling and/or downsampling to reduce the size of the image volumes used for training and testing [143, 153]. The variability in slice thickness across CT scans, combined with the anatomic extent and variability in appearance of mesothelioma tumor, precluded the development of a simple method for the subsampling of image volumes for this study. The use of downsampled lower-resolution CT volumes for training three-dimensional CNNs was not

pursued in this study due to the lack of fast, high-memory GPUs available for the training of the networks; furthermore, it is unclear to what extent gains in segmentation performance achieved with increased axial context would overcome presumed reduction in segmentation performance due to the lower resolution of the predicted tumor segmentations. Finally, axial context could be incorporated in the training of deep CNNs for the full-resolution segmentation of mesothelioma by modifying the network to accept as input the sections adjacent inferior and superior to the section for which tumor segmentation will be predicted; this technique has been applied to the task of liver segmentation [154].

4.13 Conclusion

This study implemented a deep learning-based method for the automated segmentation of mesothelioma tumor on CT scans, with the principal aim to improve the segmentation performance of the method presented in Chapter 3 on scans of patients who presented with pleural effusion. Improvement in segmentation performance when compared with the Chapter 3 method was achieved through a revision of the training set of the previous study, an extensive validation methodology, and a more varied set of training sections that exhibited tumor with pleural effusion. Significantly higher agreement with observer-provided tumor segmentations, in terms of segmentation overlap and average distance between computerized and manual tumor contours, was found when compared with the method presented in Chapter 3 on a test set of CT sections that exhibited both tumor and pleural fluid. Future work will explore the incorporation of axial context in the deep CNN-based segmentation of mesothelioma tumor on CT scans and the exclusion of collapsed lung from deep learning-predicted tumor contours.

CHAPTER 5

TEXTURE ANALYSIS FOR THE DIFFERENTIATION OF MESOTHELIOMA HISTOLOGIC SUBTYPES ON CT SCANS

5.1 Introduction

As outlined in Section 1.1, the histological subtype of mesothelioma tumor is the most significant prognostic factor for this disease, with epithelioid tumor carrying a better prognosis compared with the sarcomatoid and biphasic subtypes. The standard procedure to determine histologic subtype of mesothelioma tumor is through tumor biopsy or surgical tumor resection. A single study found a higher apparent diffusion coefficient of epithelioid mesothelioma tumor on diffusion-weighted magnetic resonance imaging (DWI-MRI) when compared with sarcomatoid tumor, which indicates the potential utility of imaging in the noninvasive assessment of mesothelioma subtype [25]. The application of MRI to mesothelioma remains limited due to complex acquisition procedures and insurance reimbursement; therefore, the development of a computerized method for the assessment of mesothelioma histology using a more widely available imaging modality is important to ensure accessibility across institutions engaged in the treatment of this aggressive disease.

Texture analysis is a quantitative method that employs mathematical features calculated from the statistical and/or morphological characteristics of images; this method has been applied to the image-based classification of tumor and prediction of patient survival in cancers of the thorax [155–157]. A single study has applied CT texture feature analysis to the differentiation of benign and malignant pleural lesions [158], and several studies have investigated the correlation of imaging findings (e.g., presence of pleural plaques, pleural effusion) on CT with histologic subtype of mesothelioma tumor [20, 159, 160]; however, no study has investigated the use of CT texture analysis for the differentiation of mesothelioma tumor subtype. The development of non-invasive, image-based computerized methods to differentiate mesothelioma histology has the potential to complement other methods of histologic

assessment for this disease and aid in the treatment planning of mesothelioma patients. Furthermore, image-based methods could provide a histologic assessment of the tumor that will not be as affected by the sampling bias of needle biopsies and other localized methods. The purpose of this study was to investigate texture analysis for the differentiation of epithelioid and "non-epithelioid" (i.e., sarcomatoid and biphasic) mesothelioma tumor on CT scans.

5.2 Patient Cohort

Patients were retrospectively collected for this study from the set of patients who underwent surgical tumor resections at the University of Chicago between January 2008 and August 2018. Sarcomatoid patients generally do not undergo surgical resections due to the poorer prognosis of this tumor subtype when compared with patients with epithelioid tumors; therefore, nine patients treated at our institution who had been diagnosed as sarcomatoid by pleural biopsy were included in the cohort initially collected for this study to supplement the number of sarcomatoid patients available for the analysis. A single study found that 20% of mesothelioma tumors diagnosed as epithelioid based on pleural biopsy were found to be of biphasic histology at extrapleural pneumonectomy (EPP) [161]; therefore, all epithelioid patients included in this study were diagnosed after surgical resection of tumor. Similarly, truly biphasic tumors could initially be misdiagnosed as sarcomatoid based on pleural biopsy; however, as biphasic and sarcomatoid tumors were grouped together as "non-epithelioid" tumors in the classification task of this study, the inclusion of sarcomatoid patients diagnosed based on pleural biopsy was deemed to be appropriate.

Talc pleurodesis has been associated with inflammation in mesothelioma patients [42]; it is unclear how these changes would affect texture on CT scans. Therefore, patients who did not have a contrast-enhanced CT scan available prior to surgical resection or prior to talc pleurodesis were excluded from this study; these criteria excluded 128 patients of the 215 patients initially collected. A further seven patients were excluded from the patient cohort due to 1) significant fluid component of tumor (as indicated by 0 HU, the pixel value of water

Table 5.1: Patient characteristics ($n = 81$).

Characteristic	All ($n = 81$)	Epithelioid ($n = 37$)	Non-Epithelioid ($n = 44$)
Sex			
Male	66 (81%)	29 (78%)	37 (84%)
Female	15 (19%)	8 (22%)	7 (16%)
Age (years)			
Median	69	68	71
Range	45–88	45–81	51–88
Histology			
Epithelioid	37 (46%)	37 (100%)	–
Biphasic	33 (41%)	–	33 (75%)
Sarcomatoid	11 (13%)	–	11 (25%)
Days between scan and surgical resection*			
Median	28	22	29
Range	1–236	3–236	1–135

* 9 non-epithelioid patients did not undergo surgical resection

on CT scans, being within one standard deviation of the mean of tumor pixel values; $n = 4$), 2) contoured tumor area less than 144 pixels (equivalent to the area of a 12×12-pixel square ROI; $n = 2$), and 3) chest wall invasion (a single sarcomatoid patient was excluded). This study was approved by the institutional review board at our institution. The final patient cohort included 81 patients and is summarized in Table 5.1. The pathologic diagnosis of the mesothelioma tumors of this study included an estimate of the relative fraction of epithelioid and sarcomatoid tumor cells of each tumor. Figure 5.1 shows the distribution of relative epithelioid tumor component of the mesothelioma patients included in this study.

5.3 Imaging

Both in-house and outside CT scans were included in this study; scans were acquired between March 2008 and February 2019. CT scans were performed on one of the following CT systems: BrightSpeed, LightSpeed 32 Pro, LightSpeed QX/i, LightSpeed VCT, Optima CT660, Revolution GSI, Revolution HD (General Electric, Milwaukee, WI), ECLOS (Hitachi, Twinsburg, OH), Brilliance 16, Brilliance 16P, Brilliance 64, iCT 256 (Philips,

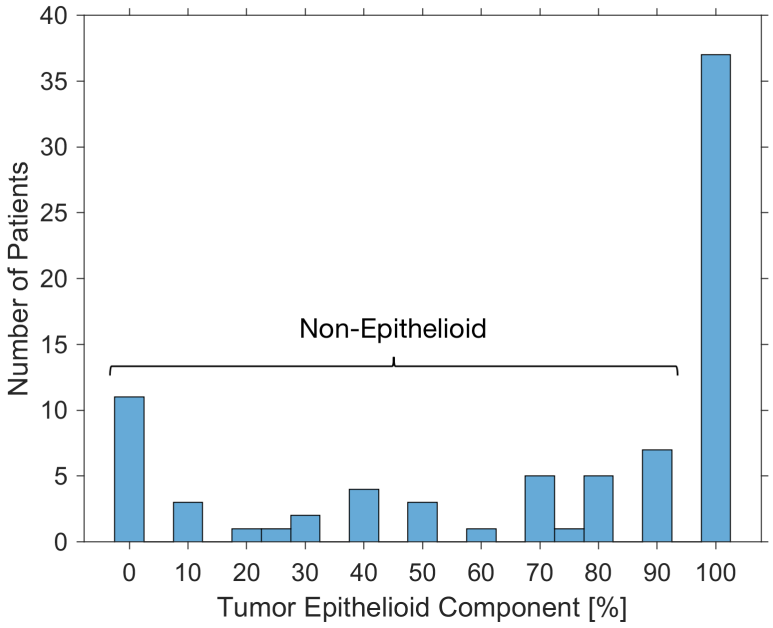

Figure 5.1: Distribution of the relative epithelioid component of mesothelioma tumors in this study ($n = 81$).

Eindhoven, Netherlands), Biograph 64, Sensation 64, SOMATOM Definition (Siemens, Munich, Germany), Aquilon, and Aquilon Prime (Toshiba, Tokyo, Japan). The reconstruction selected for each CT scan of this study corresponded to one of the "standard/nonenhancing" reconstruction kernels of each scanner manufacturer (see Table 5.2) [162].

5.4 Slice Selection and Segmentation

An imaging scientist trained in thoracic anatomy reviewed each CT scan to select the single series to maximize consistency across reconstruction kernel and contrast enhancement for scans of this study. A radiologist experienced in the measurement of mesothelioma selected the single section of each scan with the largest cross-sectional area of tumor that had a clear demarkation of tumor from 1) surrounding anatomic structures and 2) pleural effusion and atelectatic lung. This approach of selecting a single CT section with the largest cross-sectional area of tumor for analysis has been used in other studies on CT texture feature analysis [163–165]. The imaging scientist manually constructed tumor contours on the se-

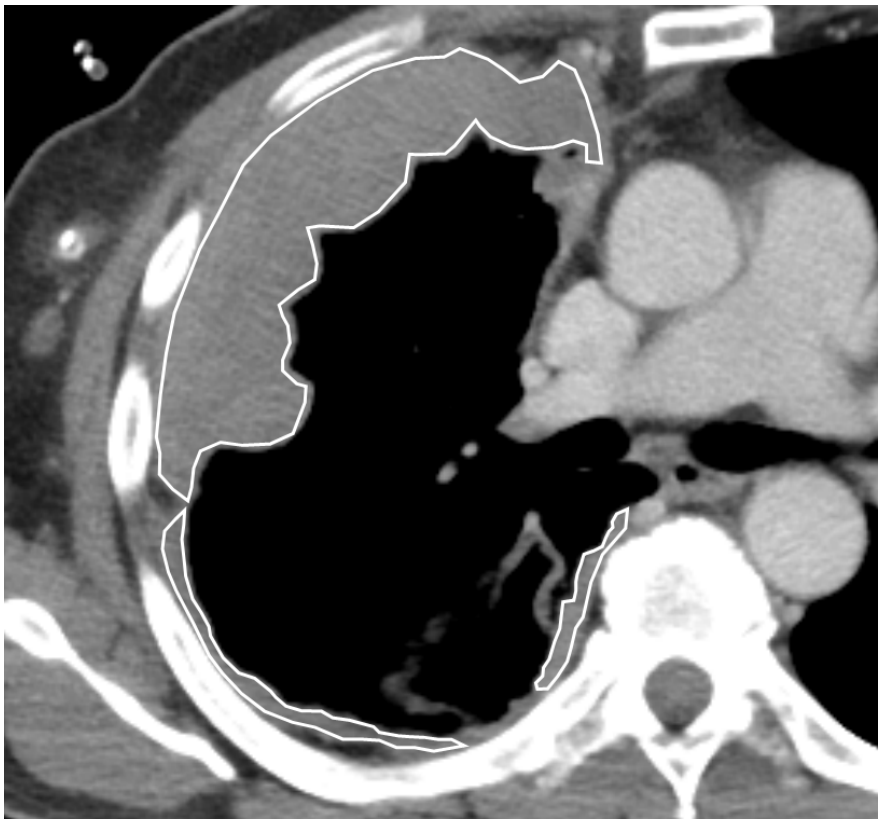

Figure 5.2: Tumor contours (white outlines) on a patient scan of this study.

lected section of each scan. All tumor contours were subsequently reviewed, and revised if needed, by a single radiologist experienced in the diagnosis and measurement of mesothelioma tumor on CT scans. Distinct areas of tumor, as visualized on the section selected for the study, were contoured as distinct tumor ROIs (these did not necessarily correspond to distinct anatomic foci of tumor). An example of tumor contours for a CT scan included in the study is shown in Fig. 5.2.

5.5 Feature Extraction

Tumor pixels for the analysis were limited to pixels with values on the range $\mu \pm 115\,\mathrm{HU}$, where μ is the median pixel value across all contoured tumor regions on the section selected for the analysis. Any pixels with values outside this range were not considered in the analysis; pixels with values within this range were subsequently binned to 64 gray levels. The range

Table 5.2: CT scan characteristics ($n = 81$).

Characteristic	All ($n = 81$)	Epithelioid ($n = 37$)	Non-Epithelioid ($n = 44$)
Pixel Size [mm]			
Median	0.74	0.74	0.75
Range	0.51–0.98	0.51–0.89	0.56–0.98
Slice Thickness [mm]			
Median	3	3	3
Range	2–5	2.5–5	2–5
kVp [kV]			
Median	120	120	120
Range	100–140	120–140	100–140
Scanner Manufacturer			
GE	14 (17%)	6 (16%)	8 (18%)
Hitachi	1 (1%)	1 (3%)	0
Philips	52 (64%)	25 (68%)	27 (61%)
Siemens	8 (10%)	1 (3%)	7 (16%)
Toshiba	6 (8%)	4 (11%)	2 (5%)
Reconstruction Kernel			
GE: Chest	3 (4%)	1 (3%)	2 (5%)
GE: Standard	11 (14%)	5 (14%)	6 (14%)
Hitachi: 32	1 (1%)	1 (3%)	0
Philips: B	51 (63%)	24 (65%)	27 (61%)
Philips: C	1 (1%)	1 (3%)	0
Siemens: B20f	1 (1%)	0	1 (2%)
Siemens: B30f	6 (8%)	1 (3%)	5 (11%)
Siemens: B31f	1 (1%)	0	1 (2%)
Toshiba: FC10	1 (1%)	1 (3%)	0
Toshiba: FC13	2 (2%)	0	2 (5%)
Toshiba: FC18	3 (4%)	3 (8%)	0

± 115 HU was selected based on the 95% limits of mesothelioma tumor pixel values on CT scans obtained by Corson et al. [18]. A similar binning approach of pixel values to the range $\mu \pm 3\sigma$, where σ is the standard deviation of the tumor ROI, with binning to 64 gray levels has been found to minimize inter-scanner variability in MRI [166] and has been used as a pre-processing step in studies on CT texture feature analysis of the chest [158, 167, 168]. Seventeen histogram-based texture features and 14 second-order texture features based on the gray-level co-occurrence matrix (GLCM) were extracted from the contoured tumor of each patient using in-house software developed in Matlab (MathWorks, Natick, MA).

Histogram-based texture features susceptible to outlier values of tumor ROIs were excluded from the analysis (this set of features included the maximum, minimum, and range of the ROI pixel values). The median of the ROI pixel values was also excluded as a histogram-based feature due to the dependence of the median pixel value of contoured tumor on the pixel-value standardization used in this study. Let G represent a collection of the pixel values of n tumor pixels contained within an ROI and let $G_{k\%}$ represent the pixel value below which $k\%$ of pixel values in G are contained. Let the probability of a given pixel value l of an ROI be represented by $p(l)$. The 17 histogram-based texture features extracted from tumor ROIs were defined as follows [96, 169, 170]:

$$\text{Mean} = \mu_G = \frac{1}{n}\sum_{i=1}^{n} G(i) \tag{5.1}$$

$$\text{Geometric mean} = \sqrt[n]{\prod_{i=1}^{n} G(i)} \tag{5.2}$$

$$\text{Mean absolute deviation} = \frac{1}{n}\sum_{i=1}^{n} |G(i) - \mu_G| \tag{5.3}$$

$$\text{Median absolute deviation} = median\left(|G(i) - median(G)|\right) \tag{5.4}$$

$$\text{Mode} = mode(G) \tag{5.5}$$

$$\text{Standard deviation} = \sigma_G = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[G(i) - \mu_G]^2} \tag{5.6}$$

$$\text{Interquartile range} = G_{75\%} - G_{25\%} \tag{5.7}$$

$$\text{Skewness} = \frac{\sqrt{n(n-1)}}{n-2} \times \frac{\frac{1}{n}\sum_{i=1}^{n}[G(i) - \mu_G]^3}{\sigma_G^3} \tag{5.8}$$

$$\begin{aligned} \text{Kurtosis} =& \frac{n-1}{(n-2)(n-3)}\left((n+1)k_1 - 3(n-1)\right) + 3\,, \\ \text{where } k_1 =& \frac{\frac{1}{n}\sum_{i=1}^{n}(G(i) - \mu_G)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(G(i) - \mu_G)^2\right)^2} \end{aligned} \tag{5.9}$$

$$\text{Energy} = \sum_{l} p(l)^2 \tag{5.10}$$

$$\text{Entropy} = -\sum_{l} p(l)\log p(l) \tag{5.11}$$

$$5\%\text{ quantile} = G_{5\%} \tag{5.12}$$

$$30\%\text{ quantile} = G_{30\%} \tag{5.13}$$

$$70\%\text{ quantile} = G_{70\%} \tag{5.14}$$

$$95\%\text{ quantile} = G_{95\%} \tag{5.15}$$

$$70\%\text{ balance} = \frac{G_{70\%} - \mu_G}{\mu_G - G_{30\%}} \tag{5.16}$$

$$95\%\text{ balance} = \frac{G_{95\%} - \mu_G}{\mu_G - G_{5\%}} \tag{5.17}$$

GLCM features were calculated from tumor ROIs to investigate differences in the spatial relationship of tumor pixel values between epithelioid and non-epithelioid tumors. The GLCM is an $N \times N$ matrix, P_N, constructed by counting how many times each pair of the N possible pixel values of a tumor ROI are located a distance d and angle θ from each other within a given ROI. For this study, the distance was set as $d = 1$ and the value of each second-order GLCM-based feature was calculated as the average value across the in-plane

directions $\theta = \{0°, 45°, 90°, 135°\}$. The following 14 GLCM-based second-order features were calculated from tumor ROIs [95, 169, 170]:

$$\text{GLCM energy} = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)^2 \tag{5.18}$$

$$\text{GLCM entropy} = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)\log\left(P(i,j)\right) \tag{5.19}$$

$$\text{Contrast} = \sum_{n=0}^{N-1} n^2 \left[\sum_{\substack{i=1 \\ |i-j|=n}}^{N}\sum_{j=1}^{N} P(i,j) \right] \tag{5.20}$$

$$\text{Inertia} = \sum_{i=1}^{N}\sum_{j=1}^{N} (i-j)^2 P(i,j) \tag{5.21}$$

$$\text{Absolute value} = \sum_{i=1}^{N}\sum_{j=1}^{N} |i-j| P(i,j) \tag{5.22}$$

$$\text{Inverse difference moment} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{1}{1+(i-j)^2} P(i,j) \tag{5.23}$$

$$\begin{aligned} \text{Sum of squares variance} &= \sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu)^2 P(i,j)\,, \\ \text{where } \mu &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j) \end{aligned} \tag{5.24}$$

$$\begin{aligned}
\text{GLCM correlation} =& \frac{1}{\sigma_x \sigma_y} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} ij P(i,j) - \mu_x \mu_y \right) , \\
\text{where } \mu_x =& \sum_{i=1}^{N} \left[i \sum_{j=1}^{N} P(i,j) \right] , \quad \mu_y = \sum_{j=1}^{N} \left[j \sum_{i=1}^{N} P(i,j) \right] , \\
\sigma_x =& \sqrt{\sum_{i=1}^{N} \left[\left(\sum_{j=1}^{N} P(i,j) \right) (i - \mu_x)^2 \right]}, \\
\text{and } \sigma_y =& \sqrt{\sum_{j=1}^{N} \left[\left(\sum_{i=1}^{N} P(i,j) \right) (j - \mu_y)^2 \right]}
\end{aligned} \tag{5.25}$$

$$\begin{aligned}
\text{Sum average} =& E_{\Sigma} = \sum_{k=2}^{2N} i P_{x+y}(k) , \\
\text{where } P_{x+y}(k) =& \sum_{\substack{i=1 \\ i+j=k}}^{N} \sum_{j=1}^{N} P(i,j)
\end{aligned} \tag{5.26}$$

$$\text{Sum entropy} = - \sum_{k=2}^{2N} P_{x+y}(k) \log \left(P_{x+y}(k) \right) \tag{5.27}$$

$$\text{Sum variance} = \sum_{k=2}^{2N} (i - E_{\Sigma})^2 P_{x+y}(k) \tag{5.28}$$

$$\begin{aligned}
\text{Difference average} =& E_{-} = \sum_{k=0}^{N-1} k P_{x-y}(k) , \\
\text{where } P_{x-y}(k) =& \sum_{\substack{i=1 \\ |i-j|=k}}^{N} \sum_{j=1}^{N} P(i,j)
\end{aligned} \tag{5.29}$$

$$\text{Difference entropy} = - \sum_{k=0}^{N-1} P_{x-y}(k) \log \left(P_{x-y}(k) \right) \tag{5.30}$$

$$\text{Difference variance} = \sum_{k=0}^{N-1} (i - E_i)^2 P_{x-y}(k) \tag{5.31}$$

To evaluate feature robustness to minor modifications in tumor contours, feature values were calculated from original tumor regions and tumor regions eroded by a single pixel; texture features with a Pearson correlation coefficient of $\rho < 0.5$ between non-eroded and eroded tumor regions were considered non-robust and excluded from the analysis. Feature values were normalized to zero mean and unit variance, separately for non-eroded and eroded regions, prior to the calculation of ρ. Tumor regions of size below the threshold of 144 pixels after erosion were excluded from this comparison.

5.6 Classifier Development

Stepwise feature selection using the built-in *stepwiseglm* function in Matlab was used to train a logistic regression model to classify tumors based on CT texture feature values as "epithelioid" and "non-epithelioid" (i.e., sarcomatoid and biphasic tumors) using a leave-one-case-out cross-validation (LOOCV) approach.

In a logistic regression model, we model the probability π of a binary outcome (in our case $\pi = 0$ for "epithelioid" tumor and $\pi = 1$ for "non-epithelioid" tumor) using the logit function as

$$\text{logit}\,(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_i \beta_i x_i\,, \tag{5.32}$$

where, in the case of the present study, x_i represents values of features selected for inclusion in the logistic regression model and β_i represent the coefficients of each of the selected features in the model [171]. Classifier models based on logistic regression produce classification scores in the range (0,1); logistic regression was selected for the classification task of this study due to its use in a previous study on classification of pleural lesions and the binary classification task of the present study [158]. Only linear terms (i.e., no higher-order interaction terms) were considered for inclusion in the logistic regression models trained for this study.

Stepwise feature selection is a method used to determine the optimal subset of features for a classification task. At each step of this method, a feature is either selected for inclusion

in the classifier model or an already-included feature is excluded from the model based on each feature's effect on the selection criterion of the procedure [172]. For this study, the deviance D of the classifier model was used as the selection criterion for stepwise feature selection:

$$D = -2\left[\log\left(L(b_1, y)\right) - \log\left(L(b_{\max}, y)\right]\right., \tag{5.33}$$

where $L(b, y)$ denotes the maximum value of the likelihood function L for a given model, y represents the observations in the data set (in our case, feature values), b_1 represents the set of parameters that have been selected for inclusion in the given model, and $b_{\max}$ represents the parameters of the maximal or full model [171]. In general, the difference ΔD of the deviances of two logistic regression models M_p and M_q follows a chi-squared distribution $\chi^2(p-q)$, where q and p represent the number of parameters in each model and $q < p$. For a given significance level α, if the value of ΔD is greater than the upper tail $100 \times \alpha\%$ point of the $\chi^2(p-q)$ distribution, we accept the new model as a better description of the data [171].

The "feature entry" phase of stepwise feature selection proceeds by calculating ΔD between the model with the subset of features already included in the model and each of the models obtained by adding one of the features not yet included in the model; each ΔD value is compared with a threshold α_{in} to determine whether to add the respective feature to the set of features included in the classifier. The significance threshold for including a feature in the subset of the classifier model was set as $\alpha_{\text{in}} = 0.10$ for this study; this value was chosen based on initial observations on a subset of the data set, which indicated that the default value $\alpha_{\text{in}} = 0.05$ might be too conservative for the relatively low number of cases included in this study. A "feature removal" phase is entered after a feature has been selected for inclusion in the classifier model, during which each of the selected features in the model is considered for exclusion from the model. During this phase, ΔD is evaluated between the current model and the model obtained by excluding each of the selected features, and if the significance reaches a given threshold α_{out}, the feature is removed from the subset of selected features [172]. For this study, $\alpha_{\text{out}} = 0.15$.

LOOCV is a validation approach for which a single case is left out of the entire set of cases, and the remaining cases are used to train a classifier; the trained classifier is subsequently applied to the single case left out of training to produce a classification score for that case. This procedure is repeated for the whole data set to produce classification scores for all cases [104]. In this study, stepwise feature selection and classification were performed separately for each training set of the LOOCV iterations to reduce dataset bias; the downside of this approach is that a potentially different set of features may be selected for classification for each of the LOOCV iterations. The stepwise feature selection was stopped when no statistically significant improvement was acquired through addition or removal of features in each LOOCV iteration; the combined number of feature addition and removal steps of each classifier was restricted to ten steps. Texture feature values were standardized to zero mean and unit variance across all cases used for the training of each classifier of this study; this standardization based on the training set of each LOOCV iteration was applied to the single case used as a test set for that iteration. To increase the number of samples available for training, distinct tumor ROIs were treated as individual samples for training each LOOCV classifier. For validation, the maximum score of the tumor ROIs of the patient scan under validation was used as the final classification score for that case.

As shown in Fig. 5.1, the "non-epithelioid" patient group included patients with a range of values for relative epithelioid tumor component. To assess the discrimination performance of the classification method for patients with predominantly sarcomatoid tumor, the distributions of classifier scores after the 81 LOOCV iterations were visually compared for 1) all epithelioid patients and all non-epithelioid patients ($n = 81$) and 2) all epithelioid patients and only non-epithelioid patients of $< 50\%$ epithelioid tumor component ($n = 59$).

5.7 Statistical Analysis

Receiver operating characteristic (ROC) analysis was used to assess classification performance for the task of differentiating epithelioid and non-epithelioid tumors based on clas-

sification scores acquired using the LOOCV approach [173, 174]. The area under the ROC curve (AUC) was calculated to assess the overall sensitivity and specificity of the classification scores; the SE and 95% CI of each obtained AUC value was estimated using ROCKIT software (`metz-roc.uchicago.edu`). The Wilcoxon rank-sum test was used to test the null hypothesis that the distributions of the total contoured tumor area of each scan were identical for epithelioid and non-epithelioid patients at a significance level of $\alpha = 0.05$.

5.8 Results

Figure 5.3 shows the distributions of the total tumor area contoured for each scan of the data set for epithelioid and non-epithelioid patients; the median (range) area of tumor contoured was 551 mm^2 (60–3,943 mm^2) across all patients, 292 mm^2 (87–3,943 mm^2) for epithelioid patients, and 708 mm^2 (60–3,434 mm^2) for non-epithelioid patients. The difference in median area of contoured tumor between epithelioid and non-epithelioid patients achieved statistical significance ($p = 0.045$).

The Pearson correlation coefficients ρ for texture features calculated from original and eroded tumor regions are listed in Table 5.3; based on these results, the features mode and 70% balance were excluded from stepwise feature selection for classifiers trained in this study. Figure 5.4 shows the ROC curve obtained with the logistic regression classifier scores in the differentiation of epithelioid and non-epithelioid tumors across all 81 LOOCV iterations; the corresponding AUC value for this classification task was 0.67 ± 0.06 (95% CI: [0.54, 0.77]). The distributions of classifier scores obtained from the 81 LOOCV iterations for (1) all epithelioid and all non-epithelioid patients and (2) for all epithelioid patients and only non-epithelioid patients with less than 50% epithelioid tumor component are shown in Fig. 5.5. Table 5.4 lists the features selected for inclusion in the logistic regression classifiers of the LOOCV iterations. Figure 5.6 shows the feature values across the tumor ROIs of epithelioid and non-epithelioid patients for the interquartile range and the GLCM correlation texture features, which were selected for classification in all 81 LOOCV iterations.

Table 5.3: Pearson correlation coefficient, ρ, for texture feature values calculated from original and eroded tumor regions. Feature values were normalized to zero mean and unit variance, separately for original and eroded regions, across all scans included in each comparison prior to the calculation of ρ.

Texture Feature	ρ
Mean	0.70
Geometric mean	0.73
Mean absolute deviation	0.96
Median absolute deviation	0.90
Mode	0.49
Standard deviation	0.96
Interquartile range	0.92
Skewness	0.86
Kurtosis	0.78
Energy	0.96
Entropy	0.96
5% quantile	0.88
30% quantile	0.76
70% quantile	0.82
95% quantile	0.94
70% balance	0.46
95% balance	0.67
GLCM energy	1.00
GLCM entropy	0.99
Contrast	0.98
Inertia	0.98
Absolute value	0.99
Inverse difference moment	1.00
Sum of squares variance	0.99
GLCM correlation	0.95
Sum average	0.99
Sum entropy	0.99
Sum variance	0.99
Difference average	0.99
Difference entropy	0.99
Difference variance	0.98

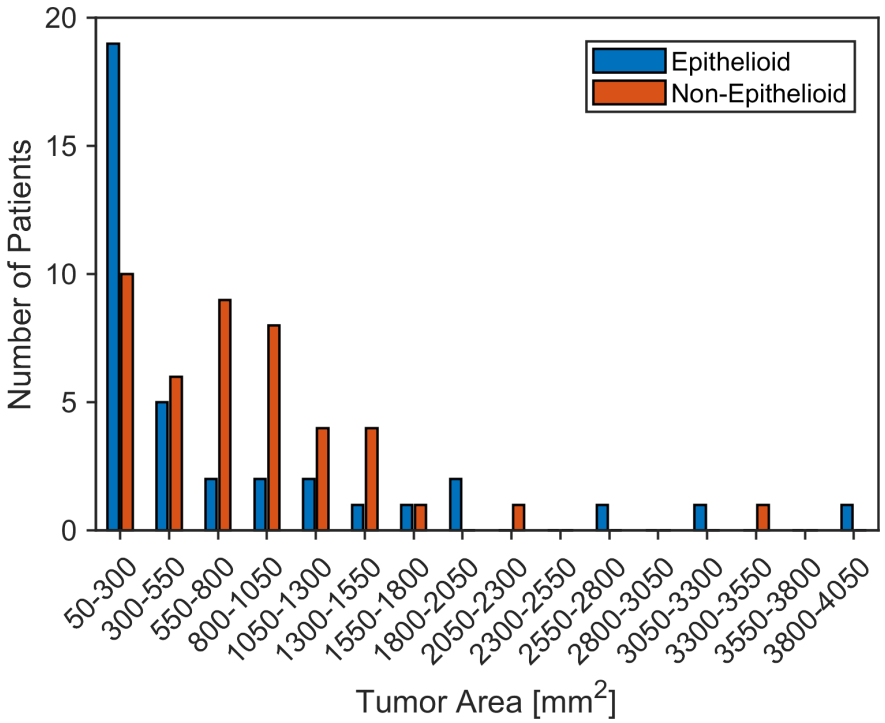

Figure 5.3: Distributions of total area of contoured tumor for the scans of this study for epithelioid and non-epithelioid patients ($n = 81$).

Table 5.4: Texture features selected for the logistic regression classifiers across the 81 LOOCV iterations of this study and the number of LOOCV iterations that each feature was selected for classification.

Texture Feature	Number of Iterations
Interquartile range	81 (100%)
GLCM correlation	81 (100%)
Median absolute deviation	80 (99%)
Inverse difference moment	74 (91%)
Energy	1 (1%)
GLCM energy	1 (1%)
Sum entropy	1 (1%)

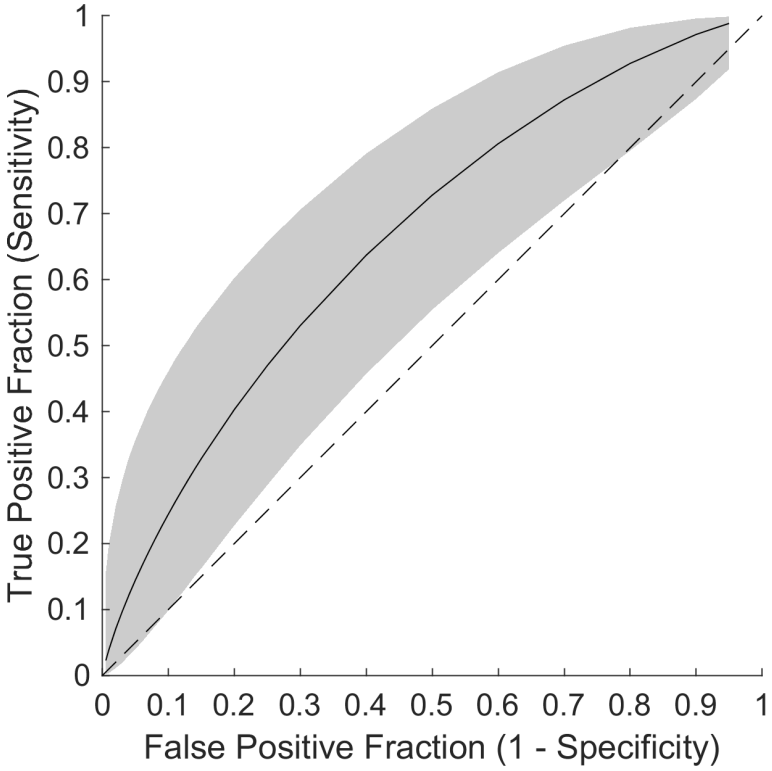

Figure 5.4: ROC curve obtained using the logistic regression classifier scores in the differentiation of epithelioid and non-epithelioid tumors for all 81 patients using a LOOCV approach; the shaded area indicates the range of the point-wise 95% confidence intervals of the ROC curve. The dashed line is the line at which the ROC curve of a classifier would indicate no discrimination capacity for the task of differentiating the two classes.

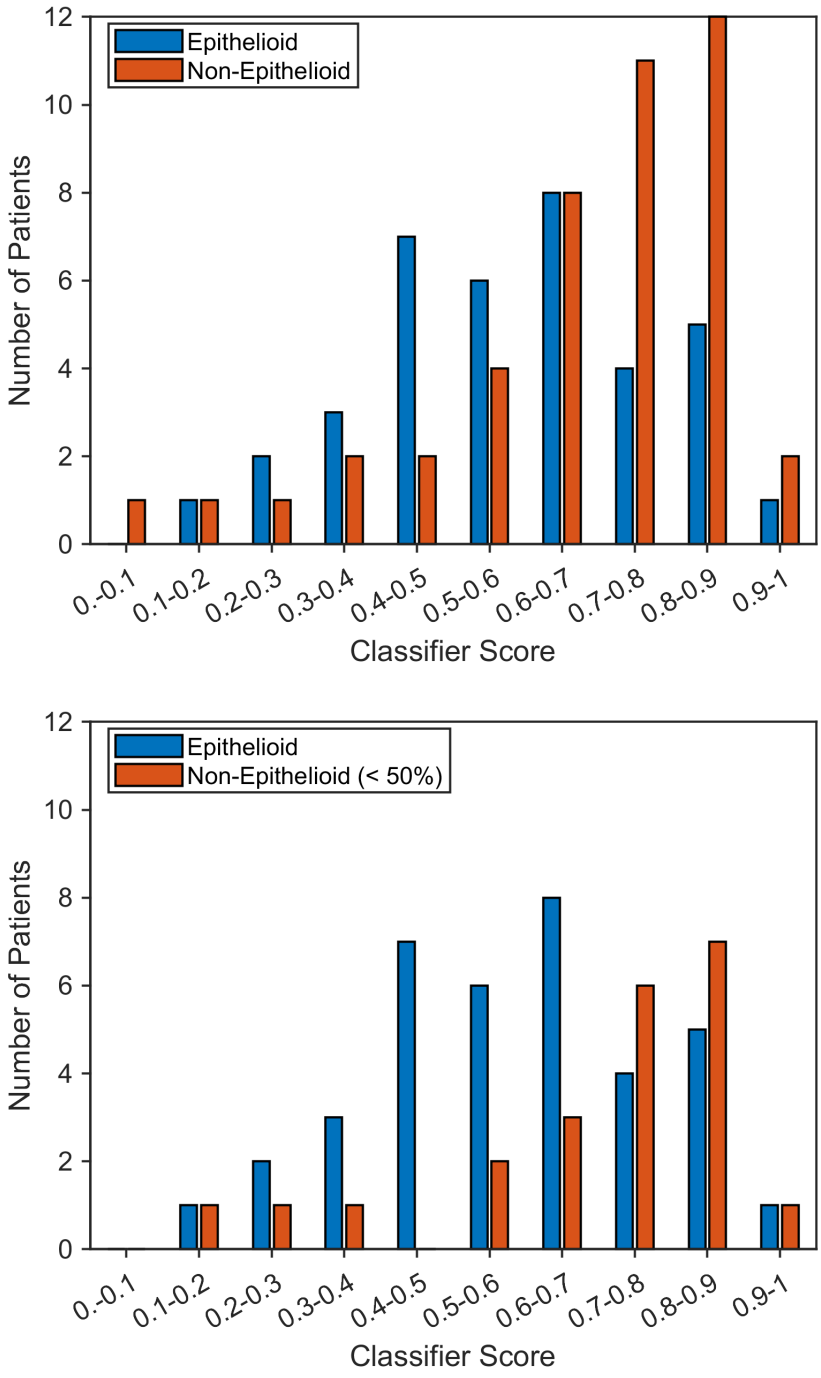

Figure 5.5: Distributions of classification scores from logistic regression classifiers trained using a LOOCV approach for the task of differentiating epithelioid and non-epithelioid tumors: (top) scores for all patients ($n = 81$), (bottom) scores for epithelioid patients and only the non-epithelioid patients with less than 50% epithelioid tumor component ($n = 59$).

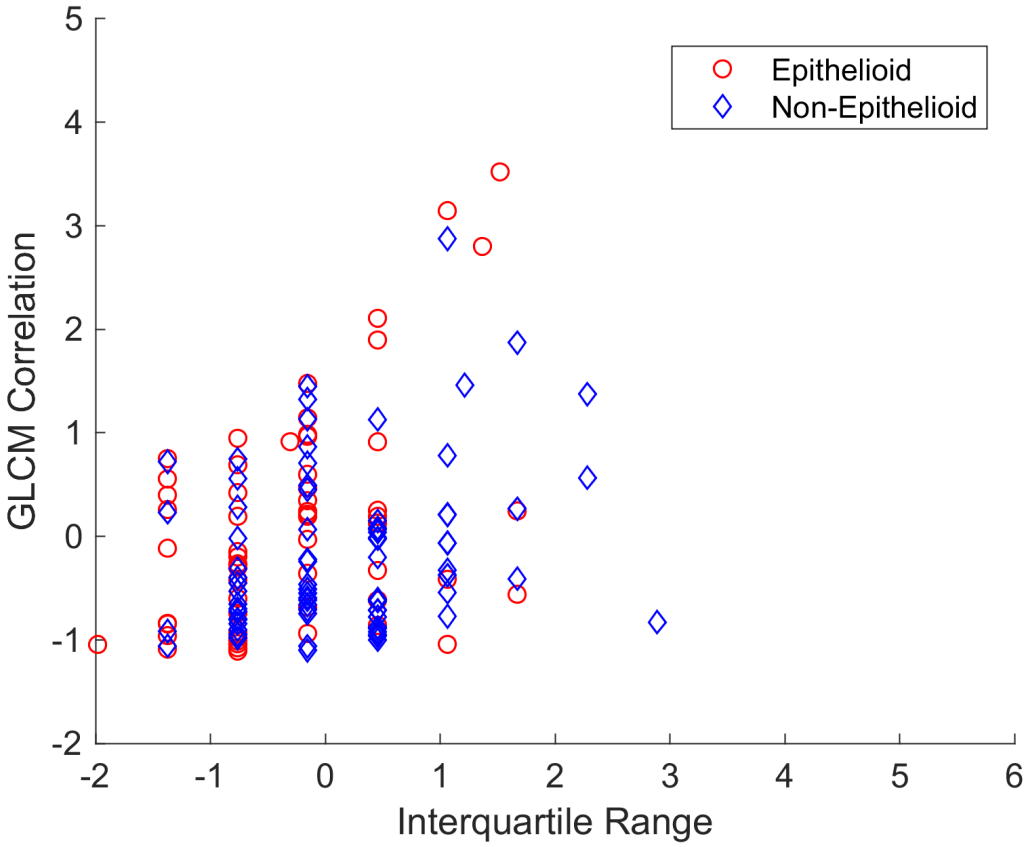

Figure 5.6: Values of the interquartile range and GLCM correlation texture features for distinct tumor ROIs of epithelioid and non-epithelioid patients. Values of each feature were standardized to zero mean and unit variance across all ROIs ($n = 138$).

5.9 Discussion

The aim of the present study was to explore the feasibility of CT texture feature analysis for the non-invasive differentiation of mesothelioma tumor histologic subtype. Logistic regression classifiers were trained using a LOOCV approach for the task of differentiating between epithelioid and "non-epithelioid" (i.e., biphasic and sarcomatoid) tumors. The AUC value calculated from the classification scores obtained in this study using a LOOCV approach across all 81 patients was 0.67 ± 0.06 with an estimated 95% CI of $[0.54, 0.77]$, which indicates an overall discrimination performance statistically significantly different from classification by chance alone (which would be indicated by an AUC value of 0.5). The limitations of this study included the relatively small patient cohort, variability in image acquisition parameters of the scans collected for this study, and differences in tumor size between the epithelioid and non-epithelioid patient groups.

The requirement that all CT scans were acquired prior to talc-pleurodesis had a large impact on the number of patient scans included in this study; this criterion excluded 128 (60%) of the 215 patients that were initially collected for this study. As outlined in Section 1.1, the majority of mesothelioma patients present with or develop pleural effusion; these patients commonly undergo talc pleurodesis, a procedure intended to fuse the pleural space to prevent fluid collection. Inflammation and pleural talc deposits are known side effects of talc pleurodesis; it is unclear to what extent these changes would affect the values of CT texture features [42]. The effect of talc pleurodesis on CT texture features should be investigated to potentially allow for the collection of a larger data set for future studies.

Due to the retrospective nature of the data collection, image acquisition parameters (see Table 5.2) varied across the scans collected for this study. Variability in CT scanner, slice thickness and in-plane pixel size have been found to influence CT texture features obtained from scans of radiologic phantoms and from scans of lung cancer patients [97, 175, 176]. As shown in Table 5.2, the pixel size and slice thickness distributions were relatively well-balanced between epithelioid patients and non-epithelioid patients; however, the distribu-

Table 5.5: The Spearman rank correlation coefficient ρ_{Spearman} between texture features that were selected in $> 50\%$ of LOOCV iterations and the area of tumor ROIs across all 81 patients.

Texture Feature	ρ_{Spearman}
Interquartile range	-7×10^{-5}
GLCM correlation	-0.93
Median absolute deviation	0.04
Inverse difference moment	0.97

tion of CT scanner manufacturer showed a greater imbalance between the epithelioid and non-epithelioid patient groups (in particular for scans acquired with Hitachi, Siemens, and Toshiba scanners). This imbalance of distributions in scanner manufacturer between epithelioid and non-epithelioid patients could have had an effect on the results obtained for the whole set of 81 patients. Future studies should limit the scans collected to one or two scanner manufacturers and aim to ensure the relatively equal distribution of scanner manufacturer between the epithelioid and non-epithelioid groups.

The comparison of contoured tumor area between epithelioid and non-epithelioid patients (see Fig. 5.3) revealed a larger median area of contoured tumor in the non-epithelioid patient group compared with epithelioid patients ($p = 0.04$). Prior studies have demonstrated the potential influence of tumor volume on CT texture features and radiomic signatures in lung cancer and head and neck cancer patients [177, 178]. Table 5.5 shows the Spearman rank correlation coefficient ρ_{Spearman} between the area of distinct tumor ROIs and the texture features that were selected in more than 50% of the LOOCV iterations across all 81 patients of this study. For two out of the four features, the absolute value of ρ_{Spearman} is greater than 0.9; this result suggests that the different distributions of tumor area between epithelioid and non-epithelioid patients could have had an effect on the classification task of this study.

The binary classification task of this study was defined in terms of "epithelioid tumor" and "non-epithelioid tumor" (which included the sarcomatoid and biphasic tumor subtypes). This binary categorization was used for this preliminary study on the texture feature-based

classification of mesothelioma tumor due to the low number of patients in the final study cohort and the significance of these categories for patient prognosis; however, histologic variation exists within the major subtypes of mesothelioma. Biphasic tumor, for example, is defined to encompass tumors of 10%–90% epithelioid component and variants of epithelioid tumor have been shown to behave similarly to sarcomatoid and biphasic tumors [10, 179]; a large multi-institutional study showed that high-grade epithelioid tumor (as based on nuclear atypia and mitotic index) might be indicative of a worse prognosis when compared with low-grade epithelioid tumor [180, 181]. Figure 5.7 shows the histologic presentation of low-grade epithelioid mesothelioma, high-grade epithelioid mesothelioma, and sarcomatoid mesothelioma. Figure 5.5 shows a comparison of the classification scores obtained from all 81 LOOCV iterations for 1) all epithelioid patients and all non-epithelioid patients and 2) all epithelioid patients and only the non-epithelioid patients with less than 50% epithelioid tumor component (i.e., predominantly sarcomatoid tumor). The separation of the distributions of classification scores for epithelioid and non-epithelioid patients shows no clear improvement when restricting the non-epithelioid group to "predominantly sarcomatoid" patients, which might be expected considering the differences in cellular characteristics between epithelioid and sarcomatoid tumors. This observation suggests that the variation in relative epithelioid and sarcomatoid tumor component within the "non-epithelioid" patient group (see Fig. 5.1) may have affected the classification results obtained in the present study. Future investigations should include a larger group of patients of predominantly sarcomatoid tumor to aid in the image-based differentiation of epithelioid and non-epithelioid mesothelioma tumor.

5.10 Conclusion

This study sought to implement a texture feature-based approach to the image-based non-invasive diagnosis of mesothelioma tumor histologic subtypes on CT scans. The tumor physiological information provided with functional imaging modalities, such as DWI-MRI, has been shown to be useful for the differentiation of epithelioid and sarcomatoid tumor;

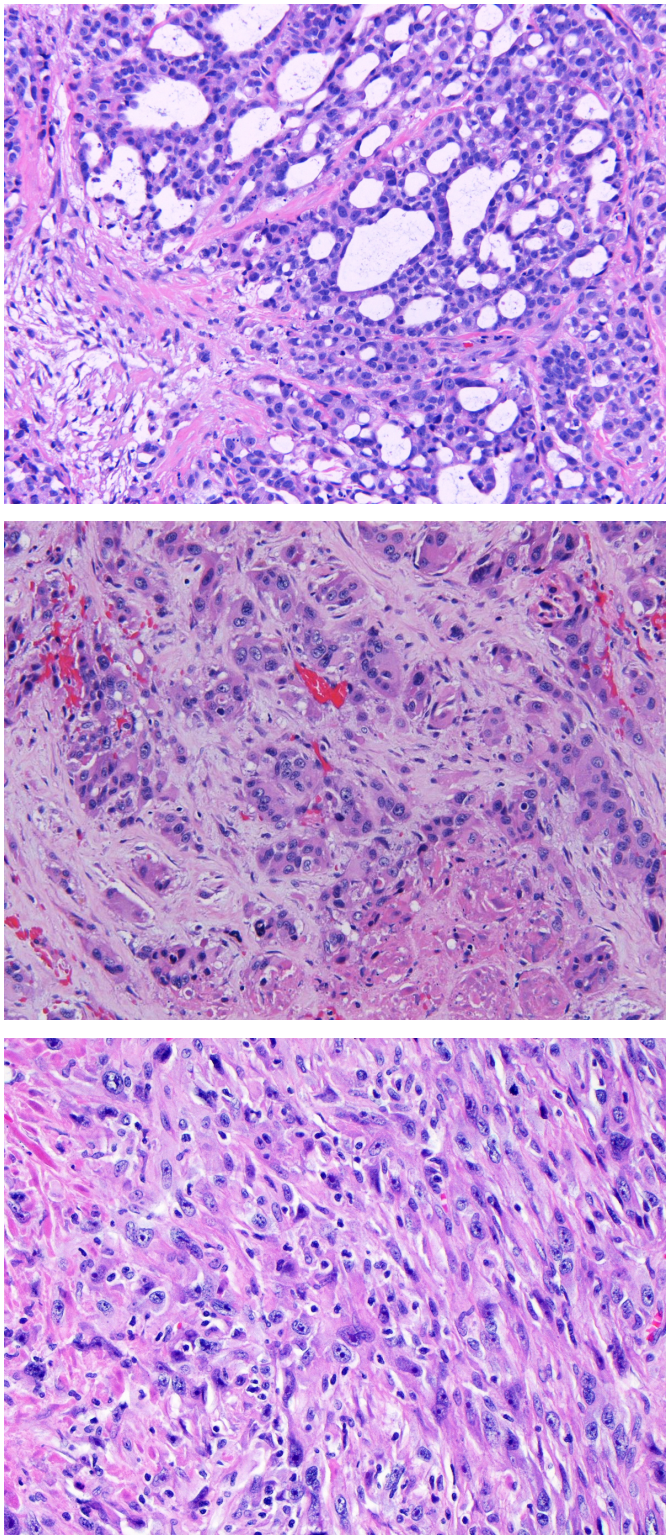

Figure 5.7: Histologic presentation of mesothelioma subtypes: (top) low-grade epithelioid tumor, (middle) high-grade epithelioid tumor, (bottom) sarcomatoid tumor. All photomicrographs were acquired with hematoxylin-eosin staining at 200× magnification. Courtesy of Dr. Aliya N. Husain.

however, the use of MRI for mesothelioma remains limited. The importance of tumor histologic subtype as a prognostic factor for mesothelioma and the wider availability of CT for the imaging of mesothelioma patients thus prompted the investigation of the present study. The ensemble of classifiers trained using the leave-one-case-out approach employed for this study showed an overall discrimination performance for the task of differentiating between epithelioid and non-epithelioid tumors that was statistically significantly different from classification by chance alone; however, the small number of patients included in the study, the heterogeneity of the data set, and imbalance of tumor size between epithelioid and non-epithelioid patients precluded any strong conclusions to be drawn from these results. Nevertheless, these results and the "radiomic pipeline" developed for this study provide motivation towards further research in this area. The requirement that scans be acquired prior to talc pleurodesis proved to be the exclusion criterion with the largest impact on patient numbers in the final cohort of this study; studies investigating the effect of talc pleurodesis on CT texture features should be undertaken to explore whether scans post-talc pleurodesis could be included in future investigations. The wide distribution of relative epithelioid tumor component of patients in the "non-epithelioid" group could have affected the classification results obtained in this study, provided that the cellular characteristics of epithelioid and sarcomatoid tumor manifest in differences detectable with CT texture features; future investigations on the image-based differentiation of mesothelioma tumor subtype on CT scans should aim to include a larger group of patients with predominantly sarcomatoid tumor.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation presented investigations on the functional, volumetric, and textural analysis of malignant pleural mesothelioma using computed tomography (CT) and deep learning, with the aims of advancing the treatment response evaluation, volumetric tumor measurement, and image-based prognostic assessment of this disease.

Chapter 2 presented an investigation on the use of dynamic contrast-enhanced CT (DCE-CT) for the assessment of mesothelioma tumor response to treatment. In this study, changes in hemodynamic parameters between scans were compared between patients on treatment and patients on observation. Differences in hemodynamic change between patient groups did not achieve statistical significance; however, the different trends observed in this study between the two patient groups indicated the potential utility of DCE-CT for the assessment of mesothelioma tumor response to treatment. The small patient cohort and the heterogeneous treatment regimens of patients in the treatment group limited the conclusions that can be drawn from this study. The investigation presented in Chapter 2 demonstrated the applicability of DCE-CT to mesothelioma and successfully incorporated dynamic CT acquisitions with standard, clinically indicated CT scans of mesothelioma patients. This study can form the basis for a larger-cohort study on DCE-CT-based treatment response assessment of patients who undergo the same treatment.

Chapter 3 presented the first exploration of the use of deep convolutional neural networks (CNNs) for the segmentation of mesothelioma tumor on CT scans. This study implemented deep CNNs of the U-Net architecture [116] for segmentation of disease separately in the left and right hemithoraces. The results of this study demonstrated the feasibility of deep learning-based segmentation of mesothelioma tumor on CT scans and showed a superior segmentation performance of the deep CNN-based method when compared with a previously published step-wise method for the automated segmentation of mesothelioma [56]. Tumor volume is a known prognostic factor of mesothelioma [50, 51, 54]; however, the lack

of time-efficient methods for the acquisition of image-based tumor volume has limited the investigation of this evaluation method for mesothelioma and has prevented the prospective validation of such a method for clinical purposes. The study presented in Chapter 3 demonstrated a state-of-the-art performance of deep CNNs for the automated segmentation of mesothelioma, which could lead to time-efficient image-based volumetric evaluations of mesothelioma patients in the clinic and in research.

Chapter 4 presented a study that improved the deep learning-based segmentation of mesothelioma tumor specifically on CT scans of patients who exhibit pleural effusion. A majority of mesothelioma patients present with or develop pleural effusion, which may complicate the segmentation of tumor due to the overlap in pixel values of mesothelioma and pleural fluid on CT scans; this aspect of the disease necessitates the validation of robust methods for the segmentation of mesothelioma on patient scans that exhibit effusion. The study in Chapter 4 achieved this goal through pre-trained convolutional layers, a comprehensive evaluation strategy, and an expanded set of CT scans enriched with cases demonstrating pleural effusion for training of the deep CNNs; a superior segmentation performance when compared with the study from Chapter 3 was demonstrated on a test set of radiologist-provided tumor segmentations on CT sections that exhibited both tumor and effusion. This investigation represents a step towards robust, generalizable and time-efficient automated segmentation of mesothelioma tumor on CT scans, which could provide clinicians with an additional data point in the management of mesothelioma patients and researchers with an efficient method for delineating tumor in CT scans that would allow for tumor-specific image-based analyses of mesothelioma.

Future research on the deep learning-based segmentation of mesothelioma tumor on CT scans should involve the acquisition of larger and more varied sets of images for the training of the deep CNNs; increasing the number of CT sections available, and the variability in disease presentation on these scans, for training deep CNNs will increase the robustness and generalizability of the deep learning-based segmentation method. To improve segmentation

performance and generalizability of the deep learning-based method presented in Chapter 4, the deep learning-based method could be applied to the segmentation of scans not included in the training set; visual observation of these deep CNN-predicted segmentations would reveal CT sections on which the performance of the method might be considered inadequate. The computer-predicted tumor contours on these sections could subsequently be manually refined and added to the training set of the method. This continual refinement of the training set of the deep CNN-based method could increase the variability of the training set of the deep CNNs without necessarily requiring full manual segmentation of tumor on these scans.

The investigations presented in Chapters 3 and 4 excluded patients who presented with tumor invasion of adjacent organs, which commonly occurs in late-stage mesothelioma patients; future studies could explore computerized methods for the segmentation of tumor in these late-stage patients. The presence of atelectatic lung (i.e., collapsed lung) on the CT scans of mesothelioma patients is a confounding factor for the segmentation of mesothelioma due to the similar visual characteristics of atelectasis and mesothelioma tumor on CT scans; future research could investigate the improvement of exclusion of collapsed lung from computerized mesothelioma tumor contours and explore the identification of atelectasis on CT scans. The analysis of an international database of mesothelioma patients found median survival for epithelioid, biphasic, and sarcomatoid patients of 19, 13, and 8 months, respectively; presumably, the earlier detection of this disease would aid in more effective treatment and the prolonging of patients' survival. Another future direction for the automated segmentation method for mesothelioma presented in Chapter 4 is to investigate the detection of mesothelioma tumor on low-dose CT scans of patients who undergo screening due to their risk of developing other cancers of the thorax, such as lung cancer.

In the study presented in Chapter 4 of this dissertation, visual observation of computerized tumor segmentations on the validation sets suggested that axial context could improve performance in the deep CNN-based segmentation of tumor, in particular in the absence of local in-plane anatomic landmarks. Three-dimensional network architectures were not

explored in the study of Chapter 4 due to the increased computational requirements for such networks when compared with two-dimensional architectures; however, future studies on deep learning-based automated segmentation of mesothelioma should investigate the effect of three-dimensional convolutional filters on segmentation performance for this task.

Chapter 5 explored the feasibility of CT texture feature analysis for the differentiation of mesothelioma tumor histologic subtype on CT scans. The histologic subtype of mesothelioma is the most significant prognostic factor of this disease; patients with the epithelioid tumor subtype have significantly improved prognosis when compared with patients with the sarcomatoid and biphasic (mixed epithelioid and sarcomatoid) tumor subtypes. The texture feature-based classification of epithelioid and non-epithelioid (i.e., biphasic and sarcomatoid) patients of this study through a leave-one-case-out cross-validation approach obtained an area under the receiver operating characteristic curve value that indicated a discrimination performance significantly different from that by chance alone. The limitations of the data set, including the small patient cohort and between-group differences in CT scanner manufacturer and tumor size, prevented strong conclusions to be drawn from the results of this study. The requirement that scans be acquired prior to talc pleurodesis was the single most important exclusion criterion of the study presented in Chapter 5; furthermore, the wide range of the relative epithelioid tumor component of patients within the "non-epithelioid" patient group of this study could have affected the differentiation capability of the texture-based classification method. Future work on this topic could explore the classification of "non-epithelioid" patients using a narrower range of epithelioid tumor component; restricting this patient group to patients who were diagnosed as predominantly sarcomatoid (i.e., $<$ 50% epithelioid tumor component) could allow for a clearer differentiation of the two patient groups through texture analysis. Future studies should investigate the effect of talc pleurodesis on texture feature values of mesothelioma tumor; understanding the impact of talc pleurodesis on CT texture analysis and being able to include scans of mesothelioma patients post-talc pleurodesis could allow for larger datasets for future analyses. Including a larger

group of patients with predominantly sarcomatoid tumor component could be advantageous to the differentiation of epithelioid and non-epithelioid mesothelioma. The time-consuming nature of the manual segmentation approach used for the construction of tumor contours for the study presented in Chapter 5 limited the analysis of tumor to a single CT section from each scan. Pathologic analysis of mesothelioma has revealed variation in tumor histologic subtype based on the sampling of tissue at pleural biopsy [161]. Future investigations on the topic of texture-based histologic classification of mesothelioma should aim to analyze larger volumetric segmentations of tumor for a more thorough sampling of the underlying tumor pathology.

This dissertation explored a variety of techniques for the extension of image-based analysis of mesothelioma for the evaluation of treatment response and patient prognosis. The first investigation of DCE-CT for treatment response evaluation of mesothelioma patients in a dedicated study revealed trends that could indicate the utility of this imaging technique for mesothelioma. Deep-learning based segmentation methods were explored for the first time for the automated segmentation of mesothelioma and achieved a significantly improved performance when compared with a previously published automated segmentation method. The generalizability of the deep CNN-based method for the segmentation of mesothelioma was improved through a thorough evaluation of the method on CT scans of patients that exhibited both tumor and pleural effusion, which commonly accompanies mesothelioma tumor; the automated segmentation of mesothelioma on CT scans that present with pleural effusion has not been explored in a dedicated study before. Finally, the feasibility of texture-based differentiation of tumor histologic subtype was investigated; the image-based classification of tumor histologic subtype, a significantly prognostic factor in mesothelioma, on CT scans has not been investigated in a dedicated study prior to the study presented in this work. Between-group differences in scan acquisition and tumor size of this study precluded strong conclusions to be drawn from the results, which should be confirmed in a future study of a larger patient cohort and less variable scan characteristics.

REFERENCES

[1] L. T. Nickell, J. P. Lichtenberger Iii, L. Khorashadi, G. F. Abbott, and B. W. Carter. Multimodality Imaging for Characterization, Classification, and Staging of Malignant Pleural Mesothelioma. *RadioGraphics*, 34(2):1692–1706 (2014).

[2] M. Ray and H. L. Kindler. Malignant Pleural Mesothelioma. *Chest*, 136(3):888–896 (2009).

[3] N. P. Campbell and H. L. Kindler. Update on malignant pleural mesothelioma. *Seminars in Respiratory and Critical Care Medicine*, 32(1):102–110 (2011).

[4] B. W. S. Robinson, A. W. Musk, and R. A. Lake. Malignant mesothelioma. *Lancet*, 366(9483):397–408 (2005).

[5] T. A. Sporn. Mineralogy of Asbestos. In A. Tannapfel, editor, *Pathology of Asbestos-Associated Diseases*, Recent Results in Cancer Research, pages 1–10. Springer, Berlin, Heidelberg (2014).

[6] J. C. Wagner, C. A. Sleggs, and P. Marchand. Diffuse Pleural Mesothelioma and Asbestos Exposure in the North Western Cape Province. *Occupational and Environmental Medicine*, 17(4):260–271 (1960).

[7] O. D. Røe and G. M. Stella. Malignant pleural mesothelioma: history, controversy and future of a manmade epidemic. *European Respiratory Review*, 24(135):115–131 (2015).

[8] D. W. Cugell and D. W. Kamp. Asbestos and the Pleura. *Chest*, 125(3):1103–1117 (2004).

[9] A. C. Bibby, et al. Malignant pleural mesothelioma: an update on investigation, diagnosis and treatment. *European Respiratory Review*, 25(142):472–486 (2016).

[10] A. N. Husain, et al. Guidelines for Pathologic Diagnosis of Malignant Mesothelioma 2017 Update of the Consensus Statement From the International Mesothelioma Interest Group. *Archives of Pathology & Laboratory Medicine*, 142(1):89–108 (2018).

[11] V. W. Rusch, et al. Initial Analysis of the International Association For the Study of Lung Cancer Mesothelioma Database. *Journal of Thoracic Oncology*, 7(11):1631–1639 (2012).

[12] F. Galateau Salle, et al. New Insights on Diagnostic Reproducibility of Biphasic Mesotheliomas: A Multi-Institutional Evaluation by the International Mesothelioma Panel From the MESOPATH Reference Center. *Journal of Thoracic Oncology*, 13(8):1189–1203 (2018).

[13] W. T. Vigneswaran, et al. Amount of Epithelioid Differentiation Is a Predictor of Survival in Malignant Pleural Mesothelioma. *The Annals of Thoracic Surgery*, 103(3):962–966 (2017).

[14] K. Hollevoet, et al. Serum Mesothelin for Diagnosing Malignant Pleural Mesothelioma: An Individual Patient Data Meta-Analysis. *Journal of Clinical Oncology*, 30(13):1541–1549 (2012).

[15] L. Righi, et al. BRCA1-Associated Protein 1 (BAP1) Immunohistochemical Expression as a Diagnostic Tool in Malignant Pleural Mesothelioma Classification: A Large Retrospective Study. *Journal of Thoracic Oncology*, 11(11):2006–2017 (2016).

[16] H. L. Kindler, et al. Treatment of Malignant Pleural Mesothelioma: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of Clinical Oncology*, 36(13):1343–1373 (2018).

[17] C. S. Ng, R. F. Munden, and H. I. Libshitz. Malignant pleural mesothelioma: The spectrum of manifestations on CT in 70 cases. *Clinical Radiology*, 54(7):415–421 (1999).

[18] N. Corson, W. F. Sensakovic, C. M. Straus, A. Starkey, and S. G. Armato III. Characterization of mesothelioma and tissues present in contrast-enhanced thoracic CT scans. *Medical Physics*, 38(2):942–947 (2011).

[19] M. Yamamuro, et al. Morphologic and functional imaging of malignant pleural mesothelioma. *European Journal of Radiology*, 64(3):356–366 (2007).

[20] J. M. Seely, E. T. Nguyen, A. M. Churg, and N. L. Müller. Malignant pleural mesothelioma: Computed tomography and correlation with histology. *European Journal of Radiology*, 70(3):485–491 (2009).

[21] S. Basu, B. Saboury, D. A. Torigian, and A. Alavi. Current Evidence Base of FDG-PET/CT Imaging in the Clinical Management of Malignant Pleural Mesothelioma: Emerging Significance of Image Segmentation and Global Disease Assessment. *Molecular Imaging and Biology*, 13(5):801–811 (2011).

[22] A. K. Nowak, et al. A Novel Prognostic Model for Malignant Mesothelioma Incorporating Quantitative FDG-PET Imaging with Clinical Parameters. *Clinical Cancer Research*, 16(8):2409–2417 (2010).

[23] A. Klabatsa, et al. The association of 18F-FDG PET/CT parameters with survival in malignant pleural mesothelioma. *European Journal of Nuclear Medicine and Molecular Imaging*, 41(2):276–282 (2014).

[24] R. T. Heelan, et al. Staging of malignant pleural mesothelioma: comparison of CT and MR imaging. *American Journal of Roentgenology*, 172(4):1039–1047 (1999).

[25] R. R. Gill, et al. Diffusion-weighted MRI of malignant pleural mesothelioma: preliminary assessment of apparent diffusion coefficient in histologic subtypes. *American Journal of Roentgenology*, 195(August):W125–W130 (2010).

[26] V. W. Rusch. A Proposed New International TNM Staging System for Malignant Pleural Mesothelioma. *Chest*, 108(4):1122–1128 (1995).

[27] V. W. Rusch, et al. The IASLC Mesothelioma Staging Project: Proposals for the M Descriptors and for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Mesothelioma. *Journal of Thoracic Oncology*, 11(12):2112–2119 (2016).

[28] A. K. Nowak, et al. The IASLC Mesothelioma Staging Project: Proposals for Revisions of the T Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Pleural Mesothelioma. *Journal of Thoracic Oncology*, 11(12):2089–2099 (2016).

[29] D. Rice, et al. The IASLC Mesothelioma Staging Project: Proposals for Revisions of the N Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Pleural Mesothelioma. *Journal of Thoracic Oncology*, 11(12):2100–2111 (2016).

[30] H. Pass, et al. The IASLC Mesothelioma Staging Project: Improving Staging of a Rare Disease Through International Participation. *Journal of Thoracic Oncology*, 11(12):2082–2088 (2016).

[31] A. Scherpereel, et al. Guidelines of the European Respiratory Society and the European Society of Thoracic Surgeons for the management of malignant pleural mesothelioma. *European Respiratory Journal*, 35(3):479–495 (2010).

[32] E. Taioli, A. S. Wolf, and R. M. Flores. Meta-Analysis of Survival After Pleurectomy Decortication Versus Extrapleural Pneumonectomy in Mesothelioma. *The Annals of Thoracic Surgery*, 99(2):472–480 (2015).

[33] A. Scherpereel, F. Wallyn, S. M. Albelda, and C. Munck. Novel therapies for malignant pleural mesothelioma. *The Lancet Oncology*, 19(3):e161–e172 (2018).

[34] N. J. Vogelzang, et al. Phase III Study of Pemetrexed in Combination With Cisplatin Versus Cisplatin Alone in Patients With Malignant Pleural Mesothelioma. *Journal of Clinical Oncology*, 21(14):2636–2644 (2003).

[35] G. L. Ceresoli, et al. Phase II Study of Pemetrexed Plus Carboplatin in Malignant Pleural Mesothelioma. *Journal of Clinical Oncology*, 24(9):1443–1448 (2006).

[36] G. Zalcman, et al. Bevacizumab for newly diagnosed pleural mesothelioma in the Mesothelioma Avastin Cisplatin Pemetrexed Study (MAPS): a randomised, controlled, open-label, phase 3 trial. *The Lancet*, 387(10026):1405–1414 (2016).

[37] A. S. Wolf and R. M. Flores. Current Treatment of Mesothelioma. *Thoracic Surgery Clinics*, 26(3):359–375 (2016).

[38] D. Rice, et al. Recommendations for Uniform Definitions of Surgical Techniques for Malignant Pleural Mesothelioma: A Consensus Report of the International Association for the Study of Lung Cancer International Staging Committee and the International Mesothelioma Interest. *Journal of Thoracic Oncology*, 6(8):1304–1312 (2011).

[39] J. Yanagawa and V. Rusch. Surgical Management of Malignant Pleural Mesothelioma. *Thoracic Surgery Clinics*, 23(1):73–87 (2013).

[40] T. Treasure, et al. Extra-pleural pneumonectomy versus no extra-pleural pneumonectomy for patients with malignant pleural mesothelioma: clinical outcomes of the Mesothelioma and Radical Surgery (MARS) randomised feasibility study. *The Lancet Oncology*, 12(8):763–772 (2011).

[41] R. Ismail-Khan, et al. Malignant Pleural Mesothelioma: A Comprehensive Review. *Cancer Control*, 13(4):255–263 (2006).

[42] J. G. Murray, E. F. Patz, J. J. Erasmus, and R. C. Gilkeson. CT appearance of the pleural space after talc pleurodesis. *American Journal of Roentgenology*, 169(1):89–91 (1997).

[43] B. H. Kwek, S. L. Aquino, and A. J. Fischman. Fluorodeoxyglucose positron emission tomography and CT after talc pleurodesis. *Chest*, 125(6):2356–60 (2004).

[44] S. Narayanaswamy, S. Kamath, and M. Williams. CT appearances of talc pleurodesis. *Clinical Radiology*, 62(3):233–237 (2007).

[45] E. T. H. Fysh, et al. Pleurodesis outcome in malignant pleural mesothelioma. *Thorax*, 68(6):594–596 (2013).

[46] M. J. Byrne and A. K. Nowak. Modified RECIST criteria for assessment of response in malignant pleural mesothelioma. *Annals of Oncology*, 15:257–260 (2004).

[47] S. G. Armato III and A. K. Nowak. Revised Modified Response Evaluation Criteria in Solid Tumors for Assessment of Response in Malignant Pleural Mesothelioma (Version 1.1). *Journal of Thoracic Oncology*, 13(7):1012–1021 (2018).

[48] P. Therasse, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. *Journal of the National Cancer Institute*, 92(3):205–216 (2000).

[49] E. A. Eisenhauer, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247 (2009).

[50] H. I. Pass, B. K. Temeck, K. Kranda, S. M. Steinberg, and I. R. Feuerstein. Preoperative Tumor Volume Is Associated With Outcome In Malignant Pleural Mesothelioma. *The Journal of Thoracic and Cardiovascular Surgery*, 115(2):310–318 (1998).

[51] F. Liu, et al. Assessment of therapy responses and prediction of survival in malignant pleural mesothelioma through computer-aided volumetric measurement on computed tomography scans. *Journal of Thoracic Oncology*, 5(6):879–84 (2010).

[52] T. Frauenfelder, et al. Volumetry: An alternative to assess therapy response for malignant pleural mesothelioma? *European Respiratory Journal*, 38(1):162–168 (2011).

[53] Z. E. Labby, et al. Disease volumes as a marker for patient response in malignant pleural mesothelioma. *Annals of Oncology*, 24(4):999–1005 (2013).

[54] R. R. Gill, et al. Epithelial malignant pleural mesothelioma after extrapleural pneumonectomy: Stratification of survival with CT-derived tumor volume. *American Journal of Roentgenology*, 198(2):359–363 (2012).

[55] V. W. Rusch, et al. A Multicenter Study of Volumetric Computed Tomography for Staging Malignant Pleural Mesothelioma. *The Annals of Thoracic Surgery*, 102(4):1059–1066 (2016).

[56] W. F. Sensakovic, et al. Computerized segmentation and measurement of malignant pleural mesothelioma. *Medical Physics*, 38(1):238–244 (2011).

[57] N. A. Obuchowski, et al. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. *Statistical methods in medical research*, 24(1):107–140 (2014).

[58] S. G. Armato III, et al. Radiologic–pathologic correlation of mesothelioma tumor volume. *Lung Cancer*, 87(3):278–282 (2015).

[59] Z. E. Labby, et al. Variability of tumor area measurements for response assessment in malignant pleural mesothelioma. *Medical Physics*, 40(8):081916 (2013).

[60] Z. E. Labby, et al. Optimization of Response Classification Criteria for Patients with Malignant Pleural Mesothelioma. *Journal of Thoracic Oncology*, 7(11):1728–1734 (2012).

[61] S. G. Armato III, et al. Measurement of mesothelioma on thoracic CT scans: A comparison of manual and computer-assisted techniques. *Medical Physics*, 31(5):1105 (2004).

[62] S. G. Armato III, A. K. Nowak, R. Francis, M. Kocherginsky, and M. Byrne. Observer Variability in Mesothelioma Tumor Thickness Measurements. *Journal of Thoracic Oncology*, 9(8):1187–1194 (2014).

[63] G. R. Oxnard, S. G. Armato III, and H. L. Kindler. Modeling of mesothelioma growth demonstrates weaknesses of current response criteria. *Lung Cancer*, 52(2):141–148 (2006).

[64] R. G. Figueiras, et al. Novel Oncologic Drugs: What They Do and How They Affect Images. *RadioGraphics*, 31(7):2059–2091 (2011).

[65] J. D. Wolchok, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clinical Cancer Research*, 15(23):7412–7420 (2009).

[66] V. L. Chiou and M. Burotto. Pseudoprogression and Immune-Related Response in Solid Tumors. *Journal of Clinical Oncology*, 33(31):3541–3543 (2015).

[67] L. Axel. Cerebral blood flow determination by rapid-sequence computed tomography: theoretical analysis. *Radiology*, 137(3):679–686 (1980).

[68] M. J. Blomley, R. Coulden, C. Bufkin, M. J. Lipton, and P. Dawson. Contrast bolus dynamic computed tomography for the measurement of solid organ perfusion. *Investigative Radiology*, 28:72–77 (1993).

[69] K. A. Miles, et al. Application of CT in the investigation of angiogenesis in oncology. *Academic Radiology*, 7(10):840–850 (2000).

[70] D. Prezzi, A. Khan, and V. Goh. Perfusion CT imaging of treatment response in oncology. *European Journal of Radiology*, 84(12):2380–2385 (2015).

[71] S. Sudarski, et al. Dynamic volume perfusion computed tomography parameters versus RECIST for the prediction of outcome in lung cancer patients treated with conventional chemotherapy. *Journal of Thoracic Oncology*, 10(1):164–71 (2015).

[72] L. Yang, et al. Computed Tomographic Perfusion Imaging for the Therapeutic Response of Chemoembolization for Hepatocellular Carcinoma. *Journal of Computer Assisted Tomography*, 36(2):226–230 (2012).

[73] K. A. Miles, et al. Current status and guidelines for the assessment of tumour vascular support with dynamic contrast-enhanced computed tomography. *European Radiology*, 22(7):1430–41 (2012).

[74] C. S. Patlak, R. G. Blasberg, and J. D. Fenstermacher. Graphical Evaluation of Blood-to-Brain Transfer Constants from Multiple-Time Uptake Data. *Journal of Cerebral Blood Flow & Metabolism*, 3(1):1–7 (1983).

[75] K. A. Miles and B. B. Kelley. CT measurements of capillary permeability within nodal masses: a potential technique for assessing the activity of lymphoma. *The British Journal of Radiology*, 70(829):74–79 (1997).

[76] K. A. Miles and M. R. Griffiths. Perfusion CT: a worthwhile enhancement? *The British Journal of Radiology*, 76(904):220–231 (2003).

[77] M. Koenig, et al. Perfusion CT of the brain: diagnostic approach for early detection of ischemic stroke. *Radiology*, 209(1):85–93 (1998).

[78] K. A. Miles. Perfusion CT for the assessment of tumour vascularity: which protocol? *The British Journal of Radiology*, 76:S36–S42 (2003).

[79] D. McPherson. Ultrasound characterization of acute myocardial ischemia by quantitative texture analysis. *Ultrasonic Imaging*, 8(4):227–240 (1986).

[80] D. Wei, et al. Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis. *Medical Physics*, 22(9):1501–1513 (1995).

[81] L. Kjær, P. Ring, C. Thomsen, and O. Henriksen. Texture Analysis in Quantitative MR Imaging. *Acta Radiologica*, 36(2):127–135 (1995).

[82] N. F. Vittitoe, J. A. Baker, and C. E. Floyd. Fractal texture analysis in computer-aided diagnosis of solitary pulmonary nodules. *Academic Radiology*, 4(2):96–101 (1997).

[83] K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt. Computer aided diagnosis of breast cancer on mammograms. *Breast Cancer*, 4(4):228–233 (1997).

[84] Y. Jiang, et al. Improving breast cancer diagnosis with computer-aided diagnosis. *Academic Radiology*, 6(1):22–33 (1999).

[85] P. Lambin, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446 (2012).

[86] M. L. Giger. Machine Learning in Medical Imaging. *Journal of the American College of Radiology*, 15(3):512–520 (2018).

[87] A. R. Cunliffe, S. G. Armato III, C. M. Straus, R. Malik, and H. Al-Hallaq. Lung texture in serial thoracic CT scans: correlation with radiologist-defined severity of acute changes following radiation therapy. *Physics in Medicine and Biology*, 59(18):5387–5398 (2014).

[88] N. Antropova, B. Q. Huynh, and M. L. Giger. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Medical Physics*, 44(10):5162–5171 (2017).

[89] J. P. B. O'Connor, et al. Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research*, 21(2):249–257 (2015).

[90] F. Davnall, et al. Assessment of tumor heterogeneity: An emerging imaging tool for clinical practice? *Insights into Imaging*, 3(6):573–589 (2012).

[91] B. Ganeshan, E. Panayiotou, K. Burnand, S. Dizdarevic, and K. A. Miles. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival. *European Radiology*, 22(4):796–802 (2012).

[92] H. J. Park, et al. Prediction of Therapeutic Response of Hepatocellular Carcinoma to Transcatheter Arterial Chemoembolization Based on Pretherapeutic Dynamic CT and Textural Findings. *American Journal of Roentgenology*, 209(4):W211–W220 (2017).

[93] L. A. Hunter, et al. NSCLC tumor shrinkage prediction using quantitative image features. *Computerized Medical Imaging and Graphics*, 49:29–36 (2016).

[94] M. G. Lubner, N. Stabo, E. J. Abel, A. Munoz Del Rio, and P. J. Pickhardt. CT textural analysis of large primary renal cell carcinomas: Pretreatment tumor heterogeneity correlates with histologic findings and clinical outcomes. *American Journal of Roentgenology*, 207(1):96–105 (2016).

[95] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6):610–621 (1973).

[96] Z. Huo, et al. Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection. *Medical physics*, 27(1):4–12 (2000).

[97] R. T. H. M. Larue, G. Defraene, D. De Ruysscher, P. Lambin, and W. van Elmpt. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *The British Journal of Radiology*, 90:20160665 (2017).

[98] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe. Automated analysis of mammographic densities. *Physics in Medicine and Biology*, 41(5):909–923 (1996).

[99] S. Katsuragawa, K. Doi, and H. MacMahon. Image feature analysis and computer-aided diagnosis in digital radiography: Detection and characterization of interstitial lung disease in digital chest radiographs. *Medical Physics*, 15(3):311–319 (1988).

[100] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes. Texture analysis of medical images. *Clinical Radiology*, 59(12):1061–1069 (2004).

[101] A. R. Cunliffe, S. G. Armato III, X. M. Fei, R. E. Tuohy, and H. Al-Hallaq. Lung texture in serial thoracic CT scans: Registration-based methods to compare anatomically matched regions. *Medical Physics*, 40(6):61906 (2013).

[102] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3):159–190 (2006).

[103] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts. Machine Learning methods for Quantitative Radiomic Biomarkers. *Scientific Reports*, 5(1):13087 (2015).

[104] R. Simon, M. D. Radmacher, K. Dobbin, and L. M. McShane. Pitfalls in the Use of DNA Microarray Data for Diagnostic and Prognostic Classification. *Journal of the National Cancer Institute*, 95(1):14–18 (2003).

[105] Q. Li and K. Doi. Analysis and minimization of overtraining effect in rule-based classifiers for computer-aided diagnosis. *Medical Physics*, 33(2):320–328 (2006).

[106] Y. Jiang, et al. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology*, 198(3):671–678 (1996).

[107] X. Huang, et al. CT-based Radiomics Signature to Discriminate High-grade From Low-grade Colorectal Adenocarcinoma. *Academic Radiology*, 25(10):1285–1297 (2018).

[108] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press (2016).

[109] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958 (2014).

[110] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536 (1986).

[111] Y. A. LeCun, L. Bottou, G. B. Orr, and K. R. Müller. Efficient backprop. In *Neural Networks: Tricks of the Trade*, pages 9–48. Springer (2012).

[112] W. Zhang, et al. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*, 21(4):517–524 (1994).

[113] S.-C. B. Lo, et al. Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, 8(7-8):1201–1214 (1995).

[114] H.-C. Shin, et al. 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).

[115] T. Kooi, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis*, 35:303–312 (2017).

[116] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention*, pages 234–241 (2015).

[117] B. Kayalibay, G. Jensen, and P. van der Smagt. CNN-based Segmentation of Medical Imaging Data. *arXiv: 1701 03056 [cs CV]* (2017).

[118] B. Q. Huynh, H. Li, and M. L. Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3) (2016).

[119] P. Lakhani and B. Sundaram. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2):574–582 (2017).

[120] J. Remon, N. Reguart, J. Corral, and P. Lianes. Malignant pleural mesothelioma: New hope in the horizon with novel therapeutic strategies. *Cancer Treatment Reviews*, 41(1):27–34 (2015).

[121] H. L. Kindler, et al. Multicenter, double-blind, placebo-controlled, randomized phase II trial of gemcitabine/cisplatin plus bevacizumab or placebo in patients with malignant mesothelioma. *Journal of Clinical Oncology*, 30(20):2509–2515 (2012).

[122] R. Thiam, et al. Optimizing the size variation threshold for the CT evaluation of response in metastatic renal cell carcinoma treated with sunitinib. *Annals of Oncology*, 21(5):936–941 (2010).

[123] M. Nishino, et al. Personalized tumor response assessment in the era of molecular medicine: cancer-specific and therapy-specific response criteria to complement pitfalls of RECIST. *American Journal of Roentgenology*, 198(4):737 (2012).

[124] E. Gudmundsson, et al. Dynamic contrast-enhanced CT for the assessment of tumour response in malignant pleural mesothelioma: a pilot study. *European Radiology*, 29(2):682–688 (2019).

[125] M. Wintermark, et al. Dynamic perfusion CT: optimizing the temporal resolution and contrast volume for calculation of perfusion CT parameters in stroke patients. *American Journal of Neuroradiology*, 25(5):720–729 (2004).

[126] J. Wang, N. Wu, M. D. Cham, and Y. Song. Tumor response in patients with advanced non–small cell lung cancer: perfusion CT evaluation of chemotherapy and radiation therapy. *American Journal of Roentgenology*, 193(4):1090–1096 (2009).

[127] J. A. Christner, J. M. Kofler, and C. H. McCollough. Estimating effective dose for CT using dose–length product compared with using organ doses: consequences of adopting International Commission on Radiological Protection Publication 103 or dual-energy scanning. *American Journal of Roentgenology*, 194(4):881–889 (2010).

[128] W. Huda, K. M. Ogden, and M. R. Khorasani. Converting Dose-Length Product to Effective Dose at CT. *Radiology*, 248(3):995–1003 (2008).

[129] B. B. Avants, et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033–44 (2011).

[130] L. Lam, S. W. Lee, and C. Y. Suen. Thinning methodologies – A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885 (1992).

[131] B. Efron and R. Tibshirani. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy: Rejoinder. *Statistical Science*, 1(1):77–77 (1986).

[132] S. Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65–70 (1979).

[133] J. Bland and D. Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2):135–160 (1999).

[134] J. Bland and D. Altman. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*, 327:307–310 (1986).

[135] M. R. Meijerink, et al. The use of perfusion CT for the evaluation of therapy combining AZD2171 with gefitinib in cancer patients. *European Radiology*, 17(7):1700–1713 (2007).

[136] E. Gudmundsson, C. M. Straus, and S. G. Armato III. Deep convolutional neural networks for the automated segmentation of malignant pleural mesothelioma on computed tomography scans. *Journal of Medical Imaging*, 5(3) (2018).

[137] L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302 (1945).

[138] A. A. Taha and A. Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1):29 (2015).

[139] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, volume 15, pages 315–323 (2011).

[140] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *3rd International Conference for Learning Representations* (2014).

[141] M. Abadi, et al. TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation*, pages 265–284 (2016).

[142] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105 (2012).

[143] Ö. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 424–432 (2016).

[144] H. R. Roth, et al. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, 66:90–99 (2018).

[145] P. Simard, D. Steinkraus, and J. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Seventh International Conference on Document Analysis and Recognition*, volume 1, pages 958–963 (2003).

[146] J. Deng, et al. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255 (2009).

[147] A. A. Bankier, et al. Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. *Radiology*, 285(2):584–600 (2017).

[148] A. S. Tsao, et al. A Practical Guide of the Southwest Oncology Group to Measure Malignant Pleural Mesothelioma Tumors by RECIST and Modified RECIST Criteria. *Journal of Thoracic Oncology*, 6(3):598–601 (2011).

[149] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv: 1409 1556 [cs CV]* (2014).

[150] E. Gudmundsson, C. M. Straus, and S. G. Armato III. Pre-trained deep convolutional neural networks for the segmentation of malignant pleural mesothelioma tumor on CT scans. In *Medical Imaging 2019: Computer-Aided Diagnosis*. SPIE (2019).

[151] M.-P. Dubuisson and A. Jain. A modified Hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition*, pages 566–568. IEEE (1994).

[152] A. Kawashima and H. I. Libshitz. Malignant pleural mesothelioma: CT manifestations in 50 cases. *American Journal of Roentgenology*, 155(5):965–969 (1990).

[153] Q. Dou, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*, 41:40–54 (2017).

[154] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, and H. Greenspan. Fully Convolutional Network for Liver Segmentation and Lesions Detection. In G. Carneiro, et al., editors, *Deep Learning and Data Labeling for Medical Applications*, pages 77–85. Springer International Publishing, Cham (2016).

[155] D. V. Fried, et al. Prognostic Value and Reproducibility of Pretreatment CT Texture Features in Stage III Non-Small Cell Lung Cancer. *International Journal of Radiation Oncology*Biology*Physics*, 90(4):834–842 (2014).

[156] T. P. Coroller, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiotherapy and Oncology*, 119(3):480–486 (2016).

[157] E. E. de Jong, et al. Applicability of a prognostic CT-based radiomic signature model trained on stage I-III non-small cell lung cancer in stage IV non-small cell lung cancer. *Lung Cancer*, 124:6–11 (2018).

[158] E. Pena, et al. Can CT and MR Shape and Textural Features Differentiate Benign Versus Malignant Pleural Lesions? *Academic Radiology*, 24(10):1277–1287 (2017).

[159] A. Senyigit, et al. Malignant Pleural Mesothelioma Caused by Environmental Exposure to Asbestos in the Southeast of Turkey: CT Findings in 117 Patients. *Respiration*, 67(6):615–622 (2000).

[160] J. G. Escalon, et al. Malignant Pleural Mesothelioma: Are There Imaging Characteristics Associated With Different Histologic Subtypes on Computed Tomography? *Journal of Computer Assisted Tomography*, 42:601–606 (2018).

[161] R. Bueno, et al. Pleural Biopsy: A Reliable Method for Determining the Diagnosis But Not Subtype in Mesothelioma. *The Annals of Thoracic Surgery*, 78(5):1774–1776 (2004).

[162] S. G. Armato III, 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).

[163] M. Ravanelli, et al. Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy. *European Radiology*, 23(12):3450–3455 (2013).

[164] X. Chen, et al. A Radiomics Signature in Preoperative Predicting Degree of Tumor Differentiation in Patients with Non–small Cell Lung Cancer. *Academic Radiology*, 25(12):1548–1555 (2018).

[165] G.-M.-Y. Zhang, et al. Can quantitative CT texture analysis be used to differentiate subtypes of renal cell carcinoma? *Clinical Radiology*, 74(4):287–294 (2019).

[166] G. Collewet, M. Strzelecki, and F. Mariette. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic Resonance Imaging*, 22(1):81–91 (2004).

[167] H. Bayanati, et al. Quantitative CT texture and shape analysis: Can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? *European Radiology*, 25(2):480–487 (2015).

[168] C. Dennie, et al. Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. *Quantitative imaging in medicine and surgery*, 6(1):6–15 (2016).

[169] T. Wagner. Texture Analysis. In *Handbook of Computer Vision and Applications, Vol. 2*, pages 275–308. Academic Press, San Diego, CA (1999).

[170] W. K. Pratt. Image Feature Extraction. In *Digital Image Processing*, pages 509–550. John Wiley & Sons, Inc., New York, NY (2002).

[171] A. J. Dobson. *An Introduction to Generalized Linear Models*. Chapman & Hall, Boca Raton, FL, 2nd edition (2002).

[172] B. Sahiner, H.-P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt. Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis. *Medical Physics*, 25(4):516–526 (1998).

[173] C. E. Metz. Basic principles of ROC analysis. *Seminars in nuclear medicine*, 8(4):283–98 (1978).

[174] L. L. Pesce and C. E. Metz. Reliable and Computationally Efficient Maximum-Likelihood Estimation of âĂIJProperâĂİ Binormal ROC Curves. *Academic Radiology*, 14(7):814–829 (2007).

[175] D. Mackin, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Investigative Radiology*, 50(11):757–765 (2015).

[176] M. Shafiq-ul Hassan, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Medical Physics*, 44(3):1050–1062 (2017).

[177] X. Fave, et al. Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. *Translational Cancer Research*, 5(4):349–363 (2016).

[178] M. L. Welch, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiotherapy and Oncology*, 130:2–9 (2019).

[179] W. Travis, E. Brambilla, A. Burke, A. Marx, and A. Nicholson, editors. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*. IARC Press, Lyon, France (2015).

[180] K. Kadota, et al. A nuclear grading system is a strong predictor of survival in epitheloid diffuse malignant pleural mesothelioma. *Modern Pathology*, 25(2):260–271 (2012).

[181] L. E. Rosen, et al. Nuclear grade and necrosis predict prognosis in malignant epithelioid pleural mesothelioma: a multi-institutional study. *Modern Pathology*, 31(4):598–606 (2018).