AUTOMATED METHODS FOR PULMONARY NODULE GROWTH RATE MEASUREMENT: EARLY COMPUTER-AIDED DIAGNOSIS OF LUNG CANCER FROM COMPUTED TOMOGRAPHY IMAGES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Artit Chinwattana Jirapatnakul

January 2013

AUTOMATED METHODS FOR PULMONARY NODULE GROWTH RATE MEASUREMENT: EARLY COMPUTER-AIDED DIAGNOSIS OF LUNG CANCER FROM COMPUTED TOMOGRAPHY IMAGES

Artit Chinwattana Jirapatnakul, Ph.D.

Cornell University 2013

Pulmonary nodules are visible as dense, opaque areas in the lung on computed tomography (CT) images and may be early indications of lung cancer. Pulmonary nodule growth rate is highly correlated with malignancy and therefore its evaluation is useful in clinical decision making. Automated methods have been developed for nodule growth rate measurements, but these methods exhibit large measurement error; reducing this error will enable radiologists to make better decisions regarding follow up and treatment, in turn improving patient outcomes. Four major aspects of pulmonary nodule measurement are addressed in this thesis.

A formal procedure for the comparative evaluation of different computer algorithms for pulmonary nodule change measurement has been developed that involves a standardized set of 50 CT image pairs and an analysis method. This procedure for the first time addresses the need to be able to quantitatively compare the performance of different methods. A study has been conducted in which developers of 18 computer methods participated and the results form a baseline with which to compare current and future algorithms.

Two different computer algorithm approaches were developed to reduce the uncertainty in growth rate measurements. The first approach, moment-based compensation (ZCOMP) was performed on segmented nodule images to address additional observed increased error in the z-direction compared to the xy-plane. By applying ZCOMP, volumetric measurement variability was reduced

from a 95% limits of agreement of (-24.0%, 18.2%) to (-12.4%, 12.7%) on zero-change nodules imaged on thin-slice scans of the same resolution. The second approach was developed to address difficult-to-segment nodules with complex shapes and attachments. Instead of explicitly segmenting the nodule from the lung parenchyma, the growth index from density method (GI_D) uses the density change in a region of interest as a surrogate growth measure. The GI_D method had much lower variation, (-11.0%, 12.3%) compared to a volumetric segmentation method, (-25.2%, 18.6%).

Finally, an automated method was developed for measuring murine pulmonary nodule growth from micro-CT scans, adapting work from methods developed for human patients. This provides improved accuracy for lesion growth measurements used in small animal pre-clinical studies. The method addresses the additional noise, lack of contrast, and poor calibration of micro-CT scans. The measured growth rate was compared to the exponential growth model, and on a dataset of six nodules with repeat scans, the method measured growth that was consistent with the model.

BIOGRAPHICAL SKETCH

Dr. Artit Jirapatnakul was born in 1983. In 2001, he graduated from Harriton High School in Rosemont, PA. He attended the Pennsylvania State University for his undergraduate education and received a Bachelor's of Science degree in Electrical Engineering with honors and high distinction in 2005.

He continued his education at Cornell University, starting a M.S./Ph.D. program in the fall of 2005. Eager to apply his knowledge of computer vision and image analysis to medical applications, he joined the Vision and Image Analysis group headed by Dr. Anthony P. Reeves. As a member of the group, he has worked on automated pulmonary nodule analysis from chest CT scans in close collaboration with Drs. David Yankelevitz and Claudia Henschke at Mount Sinai Hospital. He received a Master's of Science degree in Electrical and Computer Engineering in 2011 as part of his graduate career.

ACKNOWLEDGMENTS

As the journey to the end of my Ph.D. reaches its conclusion, I would like to take a moment to recognize everybody that made this possible.

I would like to express my heartfelt gratitude to my advisor, Dr. Anthony P. Reeves. He was always willing to discuss ideas and provide suggestions for improvements. Without his encouragement, guidance, and support, none of this would have been possible. I would also like to thank my other special committee members, Drs. Peter C. Doerschuk and Keith N. Snavely, for their time and support throughout my graduate career.

I have been fortunate to work with many collaborators over the years; I would like to thank and acknowledge Drs. David F. Yankelevitz and Claudia I. Henschke at Mount Sinai Hospital for their valuable input and feedback, Dr. Robert Weiss and Minxing Li from the College of Veterinary Medicine for their assistance with the mice and micro-CT scans, and Mark Riccio from the Cornell Micro CT Facility.

I would like to thank my friends and colleagues in the Vision and Image Analysis group, Drs. Alberto Biancardi, Andrew Browder, Brad Keller, Jaesung Lee, Jeremiah Wala, and Sergei Fotin, for always being there for me to discuss ideas, provide encouragement, and most of all, have fun. The journey would have been much more difficult without their friendship. I would also like to acknowledge other faculty members, staff, and students in the School of Electrical and Computer Engineering that have helped me along the way.

Finally, I would like to thank my family and friends for always believing in me and making me into who I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Radiologists consider the growth rate of a lesion as a major factor when deciding the suspicion for cancer and whether follow-up with the patient is necessary, and if so, what follow-up should be performed [1, 2]. The follow-up might range from having another visit in a year, a visit in a month, chemotherapy, or surgery to remove a cancer. Accurately measuring the growth of suspicious lesions is critical for improving the accuracy of diagnoses and, in turn, the quality of patient care.

While growth rate is relevant to all cancer types, lung cancer is of prime importance, since it is the leading cause of cancer death in both men and women, accounting for approximately 28% of all cancer deaths in 2012, according to the American Cancer Society [3]. The high contrast of pulmonary nodules seen in computed tomography (CT) images makes them highly suitable for quantitative imaging methods. Several methods and analyses have been developed to improve the accuracy and reliability of growth measurements of pulmonary nodules, the earliest manifestations of lung cancer, from CT scans of the lung. In this thesis, the following facets of growth measurement have been addressed:

- The creation of an analysis methodology and a related landmark image dataset to facilitate the relative performance of different pulmonary nodule measurement methods (Chapter 2).
- The development of computer algorithms to reduce the measurement error for CT scanners that exhibit anisotropic geometric distortion (Chapter 3).
- Developing density-based methods that provide useful measurements on complex nodules for which conventional geometric-based methods fail

(Chapter 4).

- Extending automated pulmonary nodule growth measurement methods to pre-clinical studies with mouse models imaged with micro-CT and validating the exponential growth rate model with this technique (Chapter 5).

The remainder of this chapter provides a background of lung cancer and outlines the issues in pulmonary lesion measurement.

1.1 Lung cancer diagnosis and monitoring

Lung cancer, a disease characterized by the uncontrolled growth and spread of abnormal cells in the lung, accounts for more cancer-related deaths than any other cancer in both men and women, more than prostate, breast, and colon cancer combined. If the cancer can be identified while it is still localized, the 5-year survival rate is 52%, compared to 16% for all stages of lung cancer; however, only 15% of lung cancers are currently detected at the early stage [3]. In a screening situation, the 10-year survival rate can be as high as 92% [4].

To diagnose lung cancer, radiologists image the patient using either chest x-ray or computed tomography (CT) in order to view structures within the lung. Two major studies have shown the effectiveness of CT screening for lung cancer. A large screening study of 31,567 asymptotic patients by the International Early Lung Cancer Action Program found lung cancer in 484 participants [4]. Of these, 412 (85%) had early stage (clinical stage I) cancer with an estimated 10-year survival rate of 88%. The 302 patients with stage I cancer who were treated via surgical resection within a month of diagnosis had an improved survival rate of 92%, while the patients that did not receive treatment died within five

years. A study sponsored by the National Cancer Institute with about 53,500 enrolled participants compared the efficacy of CT with chest X-ray and found a 20.3% reduction in deaths for those participants screened using CT compared with X-ray (354 versus 442 deaths for CT screening and x-ray screening respectively) [5].

To make a diagnosis from a CT image, radiologists examine the CT images to find nodules and then analyze the nodule's size, density, and growth rate to determine the course of action for the patient [6]. Large nodules, those greater than 25 or 30 mm, require immediate attention, usually in the form of a biopsy. Generally, small nodules less than 4 mm pose a low risk of malignancy and only require a follow-up scan in a year. For nodules of intermediate size between 5 mm and 25 mm, follow-up scans are often suggested. Based on these follow-up scans, the growth rate can be measured and nodules with a high growth rate will be referred for either additional CT scans, positron emission tomography scans, or biopsy [6, 7], though there are some differences among authors on the exact cutoff sizes, the growth threshold, and the follow-up scan interval.

In addition to diagnosing lung cancer, the growth rate of nodules may be used to monitor the response to therapy [8]. There have been attempts at developing consistent guidelines to determine whether a nodule is responding to treatment; one of the most widely used is the response evaluation criteria in solid tumors (RECIST) [9]. In the RECIST guidelines, measurable nodules are those whose minimum size is 10 mm, with a maximum 5 mm CT slice thickness—in RECIST, the size is defined as the longest diameter through the nodule on the slice where the nodule has the largest appearance. The change in size of all nodules is reported as one of the following categorical values. A complete response requires all the target nodules to disappear. Partial response to

therapy requires a 30% decrease in the sum of diameters for the target nodules, while progressive disease requires a 20% increase in the sum of diameters, but at least a 5 mm increase in size. Given a nodule that is 10 mm in diameter, partial response to therapy would require the nodule size to decrease to 7 mm in diameter, which is a decrease in volume of 66%, while an indication of progressive disease would require the nodule to increase in size to 12 mm, or a volume increase of 73%. Given a more reliable measure of size change, treatments could be evaluated more quickly for effectiveness.

1.2 Computed tomography

One of the most important developments for diagnosing lung cancer was the invention of the computed tomography (CT) scanner. CT scanners enable radiologists to view internal body structures in three dimensions through the use of an X-ray source and detector that are rotated around the body. The X-ray has the capability to penetrate matter, but the radiation intensity decreases while going through an object. The decrease in radiation is due to absorption and scattering as the X-ray travels through the object, which is represented by Beer's law of attenuation:

$$I(\eta) = I_0 e^{-\mu\eta}$$

which states that the intensity at distance η exponentially decreases from the initial intensity at distance 0, I_0, with the attenuation coefficient μ [10, Chapter 2]. The attenuation coefficient is dependent upon the density and atomic weight of the matter, which allows for the determination of density based on the observed X-ray attenuation.

As the X-ray source is rotated around the body, multiple 2D X-ray projections are acquired. In first-generation CT scanners, a single, pencil-like X-ray

beam would be emitted from the X-ray source, with a single detector placed opposite the source. The source and detector were moved synchronously, acquiring projections around the body. Modern CT scanners use a cone-shaped beam with a multi-row detector array. This enables the sampling of multiple sections (slices) of data per rotation of the source [11].

In early CT scanner designs, the source and detector would complete a full rotation around the body, after which the table supporting the patient would move, and the process would repeat until the full height (in the axial direction) of the desired region was imaged. Acquisition was performed in this manner due to the cabling connecting the source and detector, which would need to be allowed to wind and unwind. The introduction of slip-ring technology replaced the cables, allowing for continuous acquisition as the table moves, resulting in spiral sampling [11]. The X-ray source and detector are mounted on a gantry, and the ratio of the distance the table moves per rotation of the gantry to the height of the X-ray beam is referred to as the pitch. A pitch of 1.0 represents a linear motion of the table that is equivalent to the height of the X-ray beam [12].

A reconstruction algorithm takes these projections and forms a three-dimensional image which is presented as a stack of image slices. Early CT scanners with pencil-beam X-ray geometry used the inverse Radon transform to reconstruct the CT image [13], but modern CT scanners, due to the use of cone-beam geometry, multi-row detectors and spiral acquisition, use much more complex 3D algorithms that are described in detail by Buzug (2008) [14].

In each image slice, a pixel represents a small 3D volume and is typically called a "voxel". In the resulting CT images, the value of the voxel is related to the density of the tissue; CT scanners are calibrated so that, on the Hounsfield scale, a voxel value of 0 corresponds to water and -1000 to air [15]. Hounsfield

units are defined by the following expression:

$$H_{\text{tissue}} = \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}}} \times 1000$$

where μ are the linear attenuation coefficients to X-rays.

The quality of a CT scan depends upon several scanner parameters; for the automated analysis of images considered in this work, the most important parameters are radiation dose, slice thickness, pitch, field of view, and reconstruction parameters. The radiation exposure from CT is a concern due to its use of ionizing radiation, which may have harmful side effects such as cancer [16]; there is no consensus about how much radiation is a concern, only that radiation exposure should be minimized. Higher radiation doses allow for better quality images due to a higher signal to noise ratio, but this has to be balanced against the desire to limit radiation to the patient.

The slice thickness specifies the width of each section along the axial direction of the scanner, which is determined by the speed of table movement, the width of each detector, and the amount of overlap between detectors. Thinner slice thickness scans have more detail than scans with thicker slice thickness, but this comes with additional cost. The thinner slices result in more slices for the same imaged volume, resulting in additional workload for the radiologist to review more slices.

The pitch used for acquisition affects the scanning speed, radiation dose, and image quality [12]. If all other parameters are fixed, increasing the pitch will reduce the scanning time and either increase the effective slice thickness or increase the image noise.

The field of view controls the in-plane size of each voxel. In a whole-lung field of view, the entire lung is in view, resulting in an in-plane resolution of about 0.6–0.8 mm per voxel. If the radiologist knows the location of the nodule,

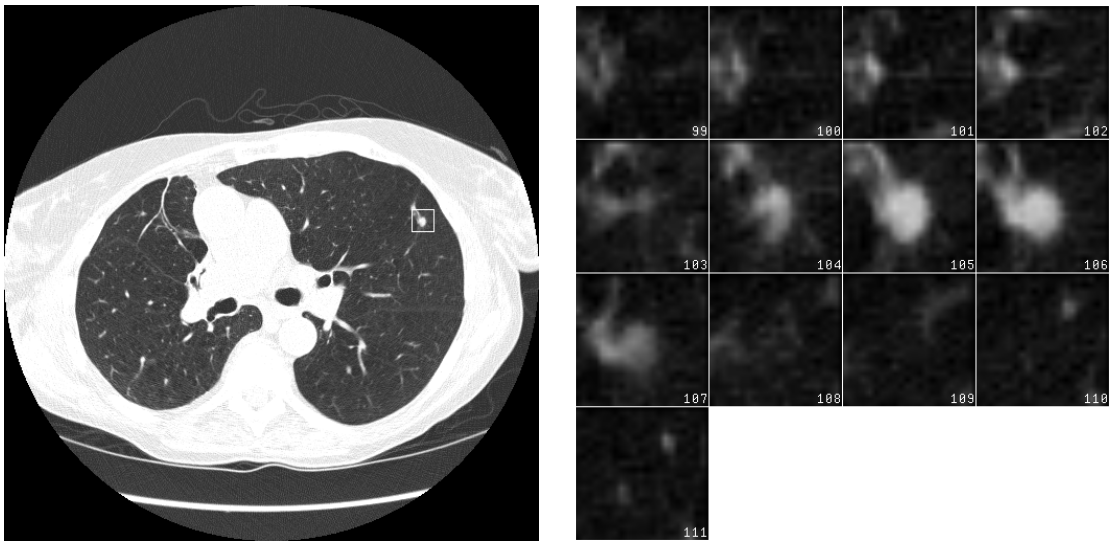
a scan with a targeted field of view can be acquired of just the region of interest. These scans typically have an in-plane resolution of 0.18 mm. Although targeted scans have a higher physical resolution, the location of the nodule needs to be known, and thus are not useful for finding new nodules.

Finally, during reconstruction, different filters may be applied to enhance the image; these filters differ primarily in the frequency-cutoff of the higher spatial frequencies. An edge-enhancing filter emphasizes high spatial frequencies, while a soft filter produces a smoother image by eliminating high spatial frequencies [17].

1.2.1 Pulmonary nodules on CT

In its earliest manifestation, lung cancer often presents as a pulmonary nodule; however, not all pulmonary nodules are malignant—some may be caused by a variety of benign conditions such as inflammation of the airways. A pulmonary nodule appears on a CT scan as a high intensity object within the lung parenchyma which does not belong to any normal anatomical structures such as vessels or airways, as shown in Figure 1.1.

Pulmonary nodules may be categorized according to their density and surrounding attachments. Nodules with a high density, called "solid nodules", have an opaque appearance on CT scans, while nodules with a low density, "non-solid nodules", have a more ill-defined appearance. Nodules with both solid and non-solid components are called "part-solid". The term "subsolid" is often used to refer to both non-solid and part-solid nodules. Examples of these nodules are shown in Figure 1.2. In addition to exhibiting different densities, nodules may either be isolated in the lung parenchyma or attached to other structures. Isolated nodules are not attached to any other high-intensity struc-



(a) (b)

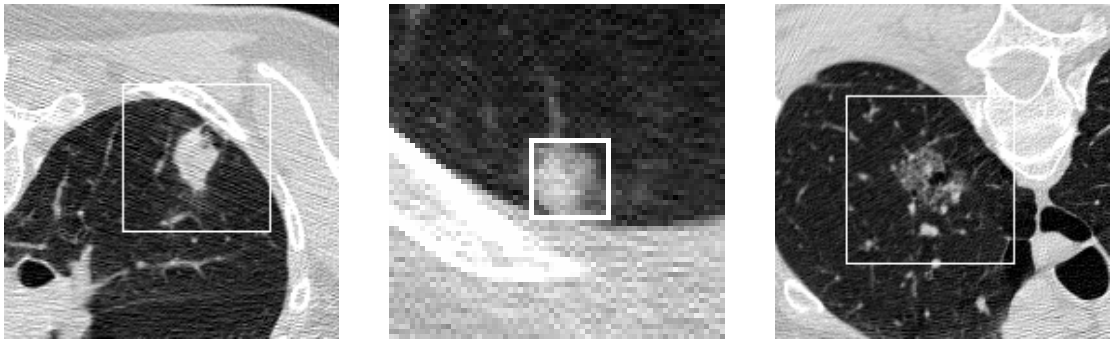
Figure 1.1: Small solid pulmonary nodule on a) single slice of a CT scan and b) several slices of a CT scan in a small region of interest

tures and are the easiest nodules to segment. Attached nodules may be attached to either blood vessels or airways, and juxtapleural nodules are attached to the chest wall. Enlarged images of the central slices of isolated, juxtapleural, and attached nodules are shown in Figure 1.3.

The majority of nodules found in a screening study are of the solid type; a study by Henschke et al. (2002) found that 88.0% (205/233) of the 233 nodules identified during baseline scans in their screening study were solid, while 12% (28/233) were non-solid nodules [18]. A majority of the malignant nodules, 82.8% (24/29), were solid or part-solid.

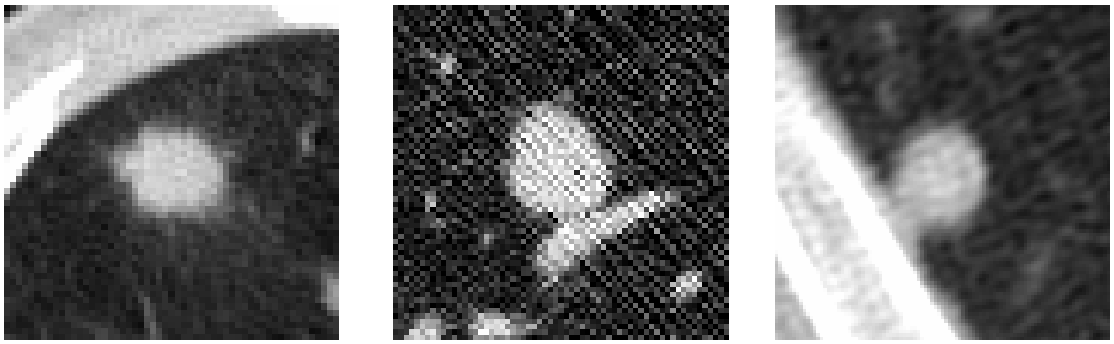
1.2.2 CT image quality issues

While CT has some unique advantages compared to other imaging modalities, it does have some flaws [19]. Due to the use of ionizing X-ray radiation, of-



(a) (b) (c)

Figure 1.2: Examples of a) solid, b) part-solid, and c) non-solid nodules on a single slice of a CT scan, with the nodule indicated by a white box. Images are shown at different magnification levels.



(a) (b) (c)

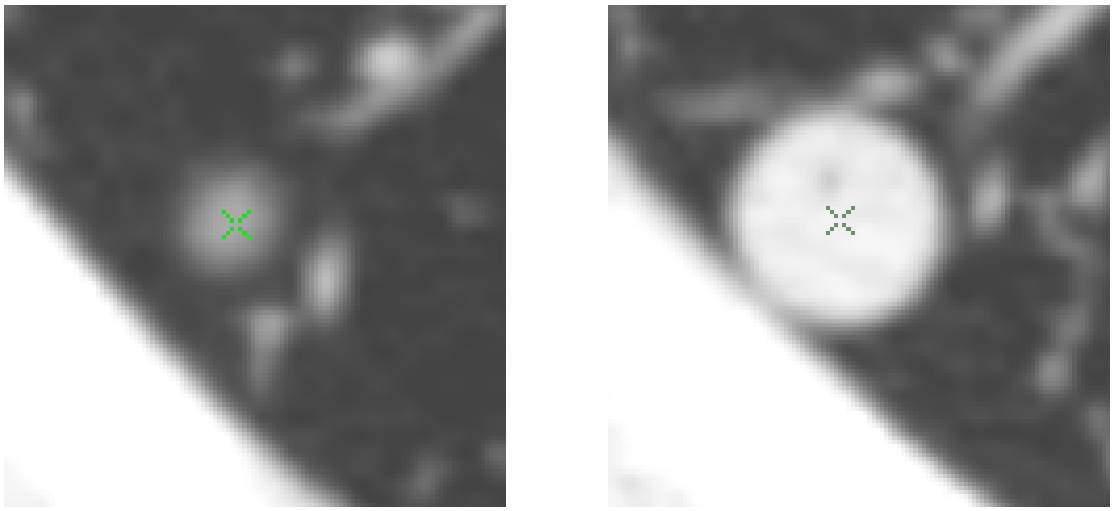
Figure 1.3: Central slices of regions containing a) isolated, b) attached, and c) juxtapleural nodules

ten image quality is comprised in order to achieve a lower radiation dose. This results in scans with additional noise compared to scans taken using a higher radiation dose. The radiation is absorbed by all tissues in the body, which results in another problem – photon starvation given the presence of a large amount of dense (absorbing) tissue to penetrate, such as bone. The tissue absorbs many of the photons, reducing the number of photons reaching the detectors resulting in an increase in noise. This noise is magnified during the reconstruction process, resulting in visible horizontal streaks in the image. Aside from tissue, the presence of metal in the body, such as pace maker wires or spinal implants, may also cause streaking artifacts due to the material's extraordinarily high density

"Beam hardening" occurs when lower energy photons in an X-ray beam are absorbed more readily as they pass through objects than higher energy photons. This effect increases the mean energy of the beam, resulting in streaking artifacts. These streaking artifacts occur between two nearby dense objects in an image. When the X-ray source is at a position where it passes through only a single object, the beam is hardened less than when the beam has to pass through both objects [19]. This occurs in bony areas of the body.

There are also artifacts caused by patient motion; if the patient moves, then structures are at different places during different portions of the acquisition, leading to misregistrations and artifacts in the reconstruction.

One of the most critical issues with respect to nodule growth measurement is the partial volume averaging effect. At the borders of objects, a voxel may not be completely filled with a single tissue type—at the edge of a pulmonary nodule, a voxel is likely to partially contain a some nodule tissue and lung parenchyma. This makes the edges of the nodule more difficult to discern, and this effect is especially pronounced on the top and bottom of the nodule, where, the voxels



(a) (b)

Figure 1.4: Example of a phantom nodule showing the edge blurring and lower intensity due to the partial voxel effect on the a) top image slice. The b) center slice is shown for comparison

on an entire slice may all be partial voxels. An example of this effect is shown in Figure 1.4. This effect can be mitigated by performing scans at a higher resolution (in-plane), such as with a targeted scan, and by using thinner image slices.

1.3 Pulmonary growth rate measurement

The growth rate of pulmonary nodules is a key indicator of malignancy. The most common method of measuring the growth rate is by measuring the size of the nodule on at least two scans taken at different times, and using the change in volume and interval between the scans to compute the growth rate [20].

The growth rate is often reported as the time it would take for the nodule to double in size, which is termed *doubling time (DT)* and given by the following

expression:

$$\mathrm{D}_T = \frac{\ln 2 \cdot \Delta t}{\ln\left(\frac{V_2}{V_1}\right)}$$

where Δt is the time interval in days between the two scans, V_2 is the nodule size measured on the second scan, and V_1 is the nodule size measured on the first scan. The doubling time expression is derived from the equation for exponential growth:

$$V_2 = V_1 \cdot e^{\lambda \cdot \Delta t} \tag{1.1}$$

$$\ln\left(\frac{V_2}{V_1}\right) = \lambda \cdot \Delta t \tag{1.2}$$

This expression can be rearranged in two ways:

$$\Delta t = \frac{\ln\left(\frac{V_2}{V_1}\right)}{\lambda} \tag{1.3}$$

$$\lambda = \frac{\ln\left(\frac{V_2}{V_1}\right)}{\Delta t} \tag{1.4}$$

To compute doubling time, the size doubles, $\frac{V_2}{V_1} = 2$, and substituting into Equation 1.3 results in:

$$\mathrm{D}_T = \frac{\ln 2}{\lambda} \tag{1.5}$$

and λ is obtained from Equation 1.4, which results in the final equation

$$\mathrm{D}_T = \frac{\ln 2 \cdot \Delta t}{\ln\left(\frac{V_2}{V_1}\right)} \tag{1.6}$$

with the parameters described above. A higher growth rate results in a smaller doubling time, and typically, a nodule with a doubling time of less than 400 days is considered to be a malignant growth rate [21]. An alternative measurement of growth rate is the *growth index (GI)*. The growth index represents the percentage increase in size per month computed using the expression

$$\mathrm{G}_I = 100 \cdot \left[\left(V_2/V_1\right)^{30.4375/\Delta t} - 1\right] \tag{1.7}$$

where V_2, V_1, and Δt are the same as in the doubling time equation. The GI expression is an exponential growth equation modified to produce a percentage result, and larger values of GI indicate faster growth. GI can be converted to DT and vice versa:

$$\mathrm{G}_I = 100\left[2^{30.4375/\mathrm{DT}} - 1\right] \tag{1.8}$$

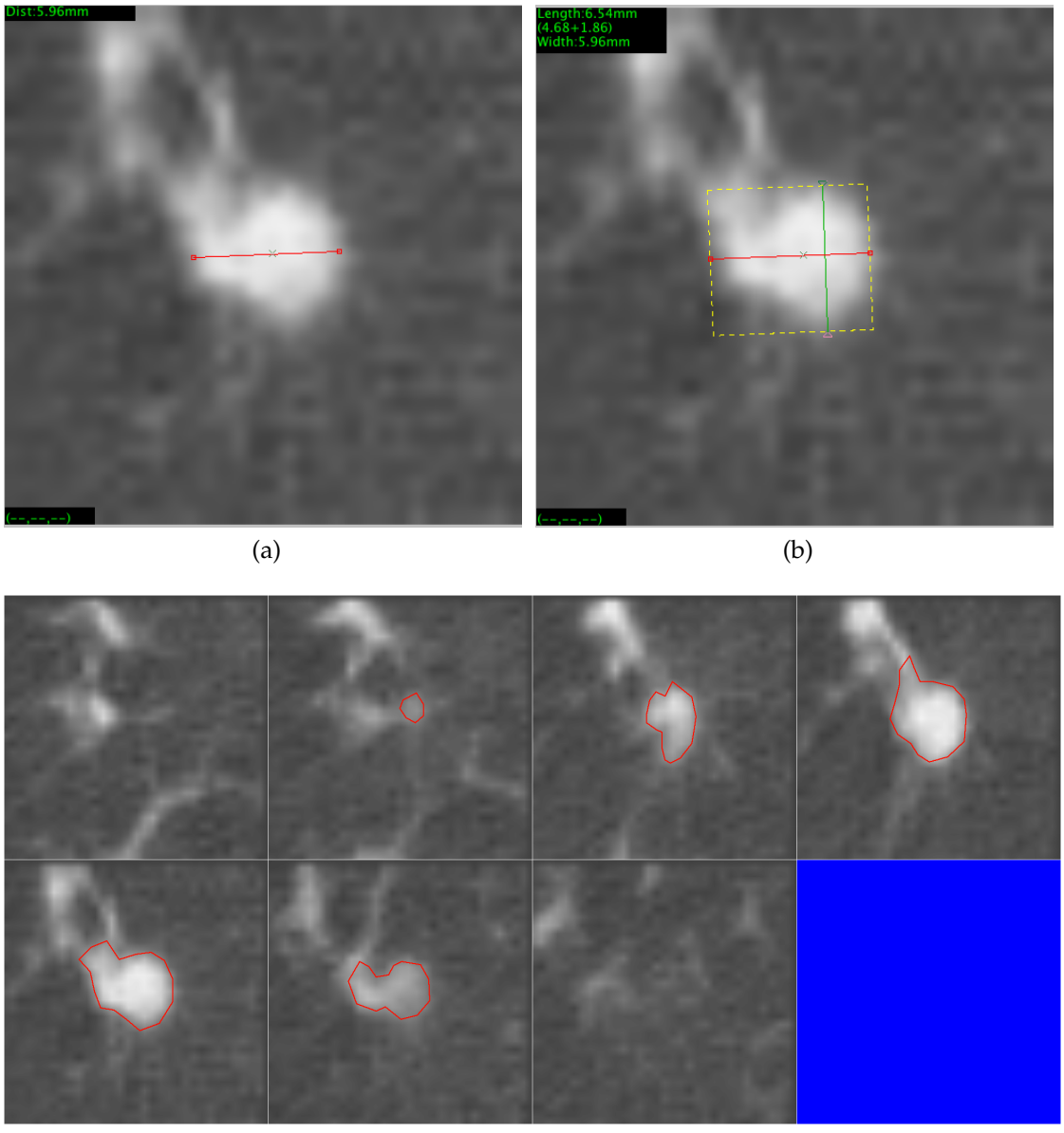
$$\mathrm{D}_T = \frac{30.4375 \cdot \ln 2}{\ln\left(\frac{\mathrm{GI}}{100} + 1\right)} \tag{1.9}$$

A doubling time of 400 days corresponds to a GI of 5.4%.

Thus, to measure the growth rate, the size change and time interval are required. The size change can be measured manually by radiologists or through automated algorithms.

1.3.1 Manual pulmonary nodule measurement

Radiologists can measure nodules using a uni-dimensional, bi-dimensional, or volumetric method. In a uni-dimensional method, the radiologist measures the size of a nodule along the largest dimension on a single central slice. For a bi-dimensional method, the radiologist marks the largest dimension on a single central slice and then indicate the largest perpendicular measurement along that dimension. Finally, in a volumetric method, the radiologist must outline the nodule on every slice on which it appears and assesses the total volume included by these boundaries. This method is the most time consuming for the radiologist of the three methods. Examples of these three methods are illustrated in Figure 1.5. Although the three methods are measuring different quantities, they are all usually converted into a volume measurement to produce a volumetric doubling time or growth index.



(a) (b)

(c)

Figure 1.5: Illustrative examples of a) uni-dimensional, b) bi-dimensional, and c) volumetric measurement methods on the same nodule. A single slice is shown for a) and b), while the all slices are shown for the volumetric measurement. a) and b) are shown with a larger magnification factor than c).

1.3.2 Automated pulmonary nodule measurement

To assist radiologists, automated methods of pulmonary nodule size measurement have been developed that are capable of measuring the volume of a nodule with minimal input from the radiologist. There are several different approaches that have been applied to this area, but the most common approach is to apply image filtering to segment the nodule from the surrounding high-intensity structures and low-intensity lung parenchyma. In this approach, a threshold is applied to the image to separate soft tissue from lung parenchyma followed by region growing or connected component analysis to eliminate structures not attached to the nodule. Finally, filtering is applied to remove noise and attached structures, such as vessels or the chest wall [22, 23]. Reeves et al. (2006) reported low volume measurement variability, 11.5%, on 50 stable nodules [23]. Some methods have explored ways to mitigate partial voxel effects. Ko et al. (2003) modeled the expected proportions of voxels belonging to the nodule and lung parenchyma, based on the size of the region and average intensity, to estimate the nodule volume [24]. Their results on phantom nodules showed reduced error compared to a fixed threshold method, with a mean absolute error of 1.6 mm^3 for the partial voxel method compared to 7.0 mm^3 for the fixed threshold method, for phantoms ranging in size from 7.5 mm^3 to 60.7 mm^3. Kuhnigk et al. (2006) applied a refinement step to a nodule segmentation method to account for the partial voxel effect by weighting the voxels in a fixed-width region at the boundary of the nodule by the voxel intensity [25]. This method with partial voxel effect compensation was found to have better volume reproducibility than a plain segmentation method on an *in vivo* dataset of 96 nodules.

In addition to image filtering-based segmentation approaches, some algorithms make use of more complex techniques such as active contours. A method

proposed by Opfer and Wiemker (2007) applied an active contour-based approach utilizing a radial basis function energy minimization to the task of nodule segmentation [26]. For 26 nodules greater than 300 mm^3, the automated method had a 95% limit of agreement of 59.8% compared to 70.2% for a radiologist. Another method by Wank, Engelmann, and Li (2007) used spiral scanning to generate a 2D image from the 3D image data followed by a method to determine the optimal outline of the nodule [27]. For 23 nodules from the LIDC database, the method achieved a segmentation overlap value of 66%.

There have also been previous attempts at measuring nodule volume and growth without explicitly segmenting the nodule. The work by Okada et al. (2005) [28] relied on the size of a Gaussian kernel determined using a multi-scale approach. The method by Kawata et al. (2005) computes nodule growth rates from the CT density histogram [29].

1.4 Evaluating pulmonary nodule growth measurement systems

When evaluating pulmonary nodules, whether to determine malignancy or asses the response to treatment, radiologists often rely on the nodule growth rate computed from CT scans. As described in the previous section, there are many methods of measuring pulmonary nodules and their growth; the critical concern is the *accuracy* and *precision* of the growth rate computed by a method.

In this context, the accuracy of a method is a measure of how close it is to the "true" growth of a nodule, while precision is a measure of the variation between measurements. If the distribution of errors is normally distributed, then the mean error would be a measure of the accuracy, while the standard deviation would be a measure of the variation or precision. An ideal method would have a mean error of 0 with no variation.

Although we are most interested in the accuracy and precision of a nodule's growth rate, most measurement methods compute growth by taking separate volume measurements on at least two scans. Given that we have accurate and precise methods of measuring time, any error or variation in the growth rate measurement would come from the volume measurements, and so, characterizing the accuracy and precision of the volume measurements is also important.

To actually determine the accuracy and precision of a method, the true growth rate of a nodule is required, but this is impossible to directly measure *in vivo*. Although comparing a method with the growth rate measured by a radiologist seems to be a natural evaluation metric, many studies, which are discussed in Section 1.4.1, have shown that radiologists' measurements are not very reliable. Another approach is to use phantoms where the true growth is known; however, these phantom nodules are often not as complex as real nodules and thus do not yield generalizable results. Yet another approach measures not the growth rate of a growing nodule, but the growth of a nodule that should not have any growth—either a stable nodule or a nodule in which repeat scans were taken during the same session. Since the true growth is known (0%), any deviations would be due to measurement error by the method, and it is hypothesized that methods with lower variations from 0% will result in better accuracy for growth rate measurement. Since the true growth of repeat scans is actually zero, they are more suited for measuring variability than stable nodules, which might in fact be growing slowly.

1.4.1 Variability of manual measurements and metrics

While growth rate is primary measurement of interest, there have been few studies that have directly determined the accuracy of the growth rate measured

by radiologists. Instead, most studies have assessed the variation of the size measurements made by radiologists. These studies generally show that radiologists' measurements show high variability and a lack of reproducibility.

There are several methods of manually measuring the size of a nodule, and a study by Jennings et al. (2004) [30] concluded that growth assessment between three different types of manual measurement—diameter, cross-sectional area, and volume computed from summing the cross-sectional areas of all the slices—often disagreed. Other studies in the area have focused on the variation of nodule size measurements, rather than growth. The variation is often characterized by interobserver agreement—how closely the measurements of multiple radiologists agree—and intraobserver agreement, which compares the measurements of the same radiologist made at multiple times or on different scans.

One study by Wormanns et al. (2000) [31] found good interobserver agreement between radiologists, based on uni-dimensional (diameter) measurements taken on CT scans. The scans were acquired using a high dose protocol (250 mA, 120 kV) and thick slice reconstruction, 5 mm and 3 mm. Other studies on size measurement have found limited agreement amongst radiologists. Bogot et al. (2005) compared bi-dimensional measurements (long-axis and short axis dimensions) made by four radiologists using film and two computer workstations [32]. On their dataset of 55 nodules, imaged using 5 mm thick scans, there was considerable variation both among the radiologists and within the radiologists – on the order of 60% in the best case. Other studies by Revel et al. (2004) [33] and Erasmus et al. (2003) [34] also found large inter- and intraobserver variation using two-dimensional and uni-dimensional measurements.

While the above studies used only uni- and bi-dimensional measurements,

there is also large variations in manual volumetric measurements, where radiologists draw a boundary around the nodule on every slice where the nodule appears. Biancardi et al. (2009) measured the variation for both manual volumetric and uni-dimensional measurements and found wide limits of agreement for the manual volumetric measurements (-23.4%, 31.8%) with even higher variation for the uni-dimensional measurements (-43.8%, 80.2%) [35]. An analysis of manual volumetric markings in the Lung Image Database Consortium study by Ross et al. (2007) [36] also found disagreement between radiologists.

Not only have studies found disagreement between radiologists, but Biancardi et al. (2007) [37] measured substantial variation in the reported nodule size depending on which metric, uni-dimensional, bi-dimensional, or volumetric, was used for nodule size measurement.

All of these studies support the conclusion that manual size measurement of pulmonary nodules is inadequate for ground truth.

1.4.2 Measurement accuracy using phantom nodules

Since manually measured nodule sizes have large variations and are not suitable for ground truth, an alternative is to use phantoms of a known size. Phantoms are objects with a similar radiodensity to pulmonary nodules placed inside another object that simulates the lung parenchyma. While phantoms may not capture all the nuances of *in vivo* nodules, they have one important advantage—phantoms have a known volume, and thus, the measurement accuracy can be determined. The FDA has been involved in the development of a database of phantom nodule scans [38], but most of the work in the area of nodule measurement variability has been performed on *in vivo* nodules, which are described in the following two sections. Phantoms are often used to help establish the er-

Table 1.1: Summary of studies on nodule volume variability on stable nodule datasets with semi-automated methods

Author	Year	# nodules	Dose	Software	Variation
Kostis [20]	2004	94	H	Research	-29.8%, 33.4% *
Reeves [23]	2006	50	H	Research	-18.3%, 18.3% †
Marchianò [42]	2009	233	L	Siemens	-27%, 27% †

* 95% limits computed from provided data
† Results did not specify the mean percentage volume change
Dose: H indicates high dose (200 mA or greater), L indicates low dose (30 mA)

ror of new pulmonary nodule measurement algorithms [39, 24], but are usually used in conjunction with experiments on *in vivo* nodules. Phantoms were also used in some of the studies described in Section 1.5.2 to determine the effect of various scanner parameters [40, 41].

1.4.3 Measurement variability using stable nodules

Stable nodules may be used to estimate the measurement variability, in addition to using phantom nodules or nodules observed in the course of normal clinical treatment. A truly stable nodule would have no change, and hence any measured change in size would be due to the measurement method or other factors unrelated to the growth of the nodule. However, actual *in vivo* stable nodules may not actually have no change, and thus the measured variations may not be reliable indicators of method variability.

Three studies have estimated the measurement variability of semi-automated nodule measurement algorithms on stable nodules with the results provided in Table 1.1. The measurement variabilities listed in the table are the 95% limits of agreement, computed according to the Bland-Altman method [43] where possible; otherwise the standard deviations were used to compute a 95% confidence

interval (indicated in the table by †). All three studies used thin-slice scans (1.25 mm or less), although the method by Reeves et al. (2006) [23] , which had the lowest variation, was evaluated on targeted scans (in-plane resolution 0.1875 mm, slice thickness 1.0 mm) which have a higher resolution than whole-lung scans (in-plane resolution of 0.6–0.8 mm). The studies by Kostis et al. (2004) [20] and Marchianò et al. (2009) [42] had similar levels of variation, on the order of $\pm$30%, although Kostis et al. used standard-dose CT scans, which have less noise than low-dose CT scans.

While these studies were able to quantify the measurement variation of semi-automated methods using stable nodules, a limitation of these studies is their assumption that stable nodules do not change in size. This means that the actual variation of the method can not be measured independently of the actual variation in the nodule size. To address this problem, some more recent studies have explored the use of zero-change scans that employ two scans acquired with a very short interval of a few minutes.

1.4.4 Measurement variability using zero-change datasets

In order to isolate the variation in measurement contributed by the system, that is, the scanner and measuring algorithm, from the variation caused by a change in the nodule, scans need to be acquired where the nodule will not change. If scans can be made of a nodule in a short period of time, on the order of minutes, then the actual change of the nodule will be zero. These types of scans are referred to as "coffee-break" or "zero-change" scans—patients are usually asked to leave and return to the CT scanner for the second scan, with enough time for a coffee break. While these scans provide insight into the variability of a measurement method, they have two disadvantages: one, the repeat scans expose

Table 1.2: Summary of studies on nodule volume variability on zero-change datasets with semi-automated methods

Author	Year	# nodules	SR (mm)	Software	Variation
Wormanns [44]	2004	151	2.2 – 20.5	Siemens	-20.4%, 21.9%
Goodman [45]	2006	43	4 – 19	GE AW	-25.6%, 25.6% *
Gietema [46]	2007	218	3.2 – 9.7	Siemens	-21.2%, 23.8%
Zhao [47]	2009	32	11 – 93	Research	-12.1%, 13.4%
Rampinelli [48]	2009	83	5 – 10	GE ALA	-38%, 60% †

(SR = Size range (diameter), AW = Advanced Workstation, ALA = Advanced Lung Analysis)

* The mean was not reported, only the width of the limits.

† Study reported results on standard-dose and low-dose scans, this represents the results on the low-dose scans.

the patient to additional, typically unnecessary radiation, and two, given that the true change is known to be zero, methods can be designed to excel at this task. Furthermore, this method does not capture the patient changes or scanner changes that occur over the much longer time interval that corresponds to clinical practice.

There are five main studies that have used zero-change scans to estimate the measurement variability of semi-automated methods [44, 45, 46, 47, 48], which are summarized in Table 1.2. Again, the variability is reported as the Bland-Altman 95% limits of agreement. Of the five studies, the two studies by Goodman et al. and Zhao et al. used standard dose scans[45, 47] while the remainder were done on low dose scans (20 – 40 mAs). The semi-automated research method developed by Zhao et al. had the lowest variation of the group, which may be due to the much larger nodules imaged on standard dose scans analyzed in their study. The remaining studies used nodule measurement algorithms available in commercial CT workstations, which had variation on the order of 25%, which is slightly better than the studies using stable nodules, except for the study by Rampinelli et al. It is not clear why their study had much

higher variation than the other studies—the authors proposed inaccurate segmentation, changes in respiratory level, and imaging at different points in the cardiac cycle as factors contributing to the variation, but these factors likely affected other studies as well.

These studies on the measurement variability of semi-automated volumetric methods using zero-change scans have shown variation on the order of 25%.

1.5 Sources of measurement variation

All nodule measurement methods have measurement variation, and in addition to the variation inherent to the method, the source of the variation may be due to either the nodule or the scan. There have been several studies that have attempted to determine what nodule characteristics and scan parameters affect the measurement variation of automated methods.

1.5.1 Nodule characteristics

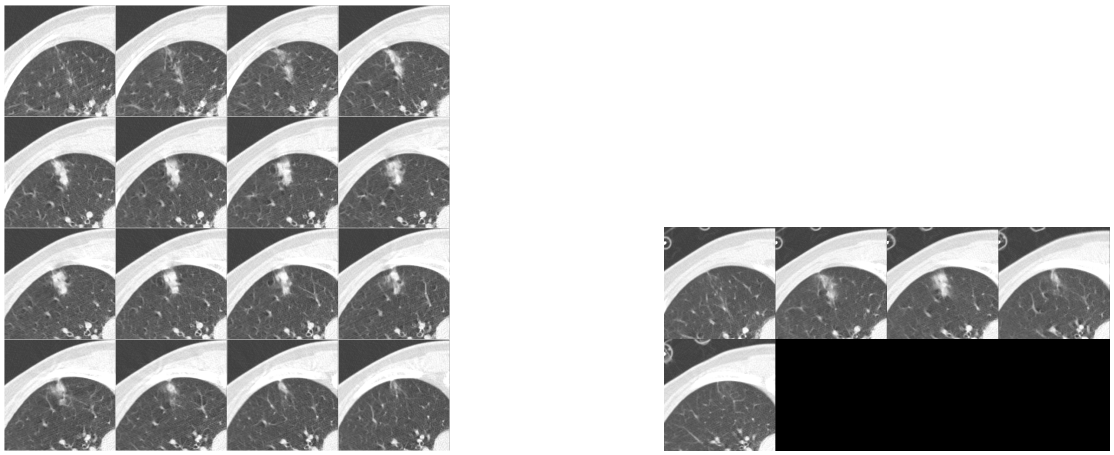
A nodule has several properties that might impact the precision of size measurement, such as its size, shape, or location. The previously mentioned study by Gietema et al. (2006) found that there was no influence of the nodule size on the measured variation, but there was a weak effect of inspiration level [46]. Their study also included what they termed "completely" segmented and "incompletely" segmented nodules, with the incompletely segmented nodules having a visually obvious exclusion of part of the nodule; the completely segmented nodules had smaller variation than the incompletely segmented nodules. In the study by Wang et al. (2008), 4225 nodules were measured by two radiologists using semi-automated software [49]. The measurement variability was found to

be affected by the nodule morphology, location, and size, however, in the study, a large majority (86%) of nodules were measured to have the same volume by both radiologists due to the exclusion of nodules where manual editing was performed by any of the radiologists. Petrou et al. (2007) found a significant influence of spiculated margins on the measurement variability [50]. Even the presence of attached structures, such as blood vessels, may impact the volume estimation [51].

1.5.2 Scan parameters

The slice thickness of the scan was found to have a significant effect on the variation by Zhao et al. (2005) [52], Petrou et al. (2007) [50], and Ravenel et al. (2008) [53], leading to most of them recommending that 1) a consistent slice thickness be used and 2) the thinnest slices possible should be acquired. An example of scans of a pulmonary nodule acquired at different slice thicknesses is shown in Figure 1.6. The results of a study by Ko et al. showed that scans with a higher current or high-frequency reconstruction algorithm had improved precision [24].

The scanner technology was also shown to affect the measurement variability. Das et al. (2007) used phantoms to quantify the absolute percentage error of four different scanner types: a single-slice spiral CT, 4-slice multidetector CT (MDCT), a 16-slice MDCT, and a 64-slice MDCT [40]. Standard and low dose protocols using thick and thin collimation were performed as well. There were statistically significant influence on the measurement variability from the scanner type, protocol, location, and nodule size. Das et al. (2007) also performed a study to determine whether the error from scanners of different manufacturers



(a) 1.25 mm (b) 2.5 mm

Figure 1.6: Example of scans of the same nodule acquired with different slice thickness

would be similar [41]. Using nodule phantoms, the 16-slice scanners from the different manufacturers had different, but similar amounts of measurement error: the mean absolute percentage error was 8.4%, 14.3%, 9.7%, and 7.5% for the Siemens, GE, Philips, and Toshiba scanners respectively.

1.6 Previous work on nodule segmentation by Reeves et al.

Most pulmonary nodule measurement algorithms measure the volume of a nodule by segmenting the nodule from the lung parenchyma and other attached structures and computing the volume of the segmented image. One such algorithm was developed by Reeves et al. (2006) [23] and was used both as a reference algorithm and as a basis for many of the improved algorithms described in the following chapters. The basic steps of the algorithm are:

1. Estimate the size and location of the nodule based on a seed point,
2. Perform resampling, applies a threshold, and
3. Filter the image to remove attached structures.

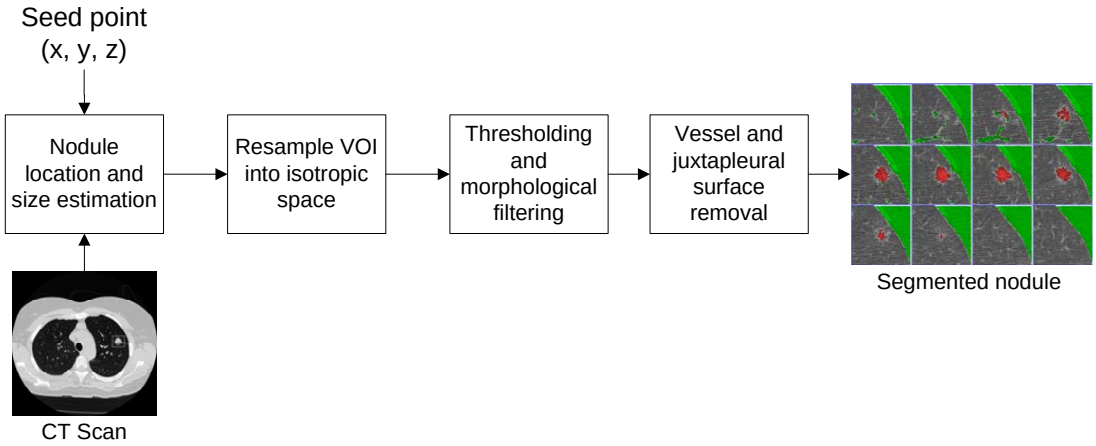


Figure 1.7: Overview of nodule segmentation method

An overview of the method is shown in Figure 1.7.

The segmentation method is initialized with a manually specified seed point within the nodule. To help reduce the variation in segmentation due to different seed point positions in the same nodule, a new estimated nodule location is determined from the seed point. Starting from the seed point, an iterative process estimates the nodule size and location using spherical Gaussian and Laplacian of Gaussian templates. An estimated location is determined by finding the location that best matches a Laplacian of Gaussian template of a given radius. Next, the location is fixed, and Gaussian templates of various sizes are tested to determine the best estimate of size for the location. This process is repeated until the location and size estimates converge or a given number of iterations is reached.

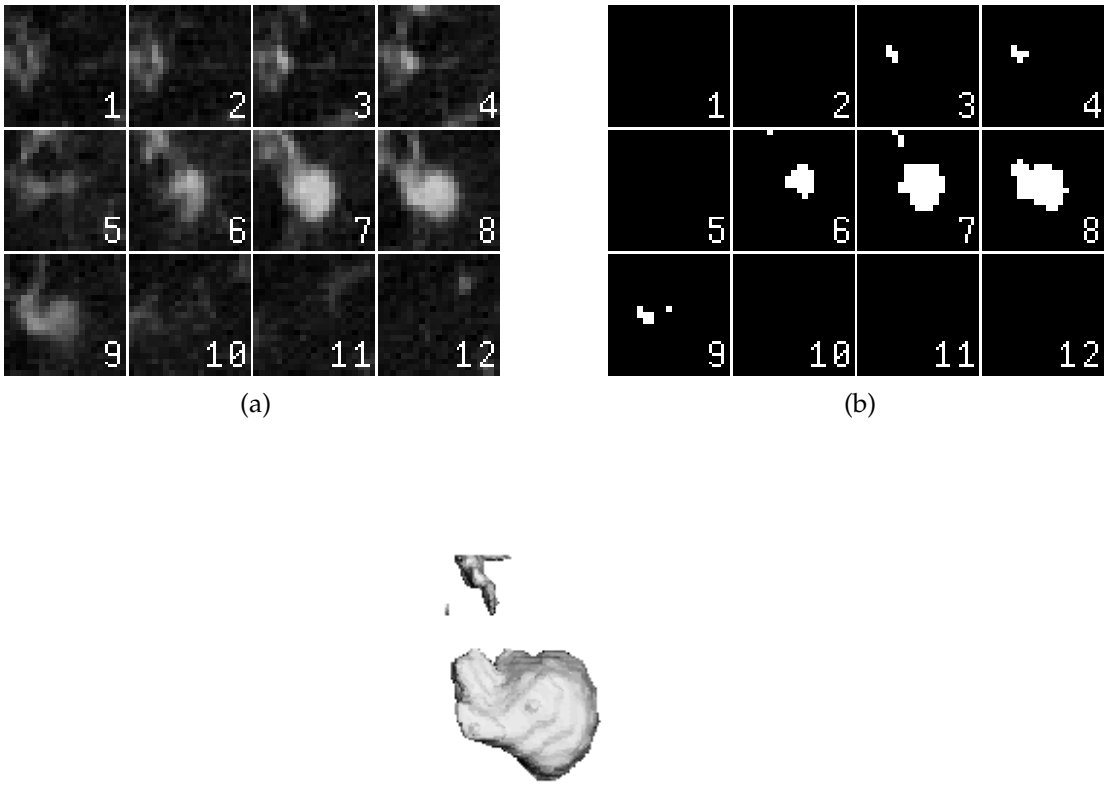
The size and location estimate are used to extract a small region of interest (ROI) from the CT scan. By selecting an ROI from the CT scan, the amount of image data to be analyzed is considerably reduced. The ROI is centered at the nodule location and has a size of twice the nodule size. The ROI is resampled into isotropic space (where the three-dimensions of each voxel are the same) using tri-linear interpolation. Typically, CT scans have a resolution of 0.6 mm x 0.6

mm x 1.25 mm, which is resampled to 0.25 mm isotropic resolution. Resampling the ROI serves two purposes: it enables subvoxel precision for segmenting the nodule, and it simplifies later processing of the image.

At this point, the algorithm has an estimate of the nodule location and size, and a resampled, isotropic ROI. The nodule is separated from the lung parenchyma by applying a threshold to the ROI; this is based on the fact that, within the lung, there are two main tissue types—-lung parenchyma and soft tissue (nodules, vessels, chest wall, blood). After applying a threshold to the ROI, the result is a binary image, with voxels belonging to soft tissue assigned a value of 1 and everything else assigned a value of 0. An example of the ROI before and after thresholding is shown in Figure 1.8.

After thresholding the image, a series of filtering operations is performed to remove noise and attached structures. A connected component analysis removes structures with a high intensity not connected to the central component. After this step, a single connected component remains – this usually consists of the nodule and adjacent structures such as blood vessels or the chest wall. To remove blood vessels, a series of morphological operations is performed. First, a morphological opening operation is applied to the image, using a spherical kernel set empirically to 0.75 times the estimated nodule radius. The opening operation removes some of the fine surface features on the nodule, so to regain these features, iterative morphological dilation is performed, starting from half the initial kernel size and halving the size in each iteration. The result of this algorithm for the nodule from Figure 1.8 is shown in Figure 1.9.

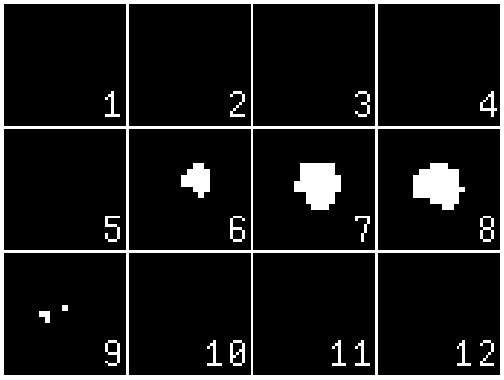
Finally, if the nodule is juxtapleural, a surface removing algorithm is performed. The details of the method are further described by Jirapatnakul et al.



(a) (b)

(c)

Figure 1.8: Example of axial slices of a nodule ROI a) before and b) after thresholding. A 3D rendering of thresholded image is shown in c).



(a)

(b)

Figure 1.9: Result of morphological filtering to remove noise and attached structures shown as a) axial slices and b) a 3D rendering

(2011) [54], but in brief, the pleural surface is modeled using a three-dimensional cubic function. Robust statistics are used to ensure that the points used in the surface estimation actually belong to the pleural surface and not the nodule or other attached structures. From these points, an estimation of the pleural surface can be computed and used to separate the juxtapleural nodule from the pleural surface.

At the end of this step, the algorithm returns a binary, isotropic image of the nodule segmentation. To compute the volume of the nodule, the number of (supersampled) voxels in the segmentation that have a value of one is summed and multiplied by the voxel size.

1.7 Overview of work on automated methods of pulmonary nodule growth measurement

The work described in this dissertation is focused on advancing the field of automated pulmonary nodule growth measurement in three main areas: determining a quantitative evaluation method through the development of a common dataset and evaluation metrics, improving pulmonary nodule growth measurement accuracy and variability, and adapting these automated methods to preclinical studies with mouse models.

The evaluation of pulmonary nodule growth measurement methods is difficult without a common dataset or evaluation criteria. In Chapter 2, the VOLumetric Change Analysis of NOdules (VOLCANO) study is described. A dataset of 49 nodules, including nodules with both zero-change and actual growth, and a single phantom was prepared and made publicly available for different research groups and manufacturers to analyze with their methods. The results of 18 different methods were evaluated for agreement using non-parametric statistical measures. This was the first study and dataset that concentrated specifically on pulmonary nodule growth measurement, and this study determined a benchmark for the state of the art.

The second major topic addressed in this dissertation is the development of two new methods for improving nodule growth rate measurement. The first method, described in Chapter 3, reduced measurement variation by applying compensation in the z-axis using moment analysis of a segmented nodule image. This addresses the uncertainty from the anisotropic CT scan acquisition method for voxels in the xy plane compared with the z plane. This method shows a promising reduction in measurement variability compared to methods

without compensation. The second method, discussed in Chapter 4, addresses nodules that are difficult to segment—nodules with complex shapes and attachments or non-homogenous density. Instead of explicitly segmenting the nodule, this method computes the change in density over a region of interest as a surrogate measure of growth. The growth rates computed by this density-based method were better able to distinguish malignant from benign nodules than the growth rates of a segmentation-based algorithm.

While these methods and many others have been developed for measuring nodule growth in human patients, there are few automated methods for measuring pulmonary nodule growth in mouse models. Many pre-clinical studies use mouse models to aid in the development of new treatments or better understand the underlying physiology of cancer. In Chapter 5, a automated segmentation algorithm is modified for measuring pulmonary nodules of small animals imaged with micro-CT. To evaluate the algorithm, instead of using manual measurements as ground truth, the growth rate measured by the algorithm was compared to the exponential growth model. On a dataset of six nodules identified on four mice, the algorithm was able to successfully segment the nodules and estimate growth rates that were good fits to the exponential model.

A final summary of all the research accomplishments is given in Chapter 6.

CHAPTER 2

EVALUATION OF AUTOMATED GROWTH MEASUREMENT METHODS: VOLCANO

A primary indicator of malignancy for early stage lung caner is the growth rate of pulmonary nodules. Accurately measuring the growth rate of pulmonary nodules is especially important in cases of malignant, slowly growing nodules—by reducing the uncertainty in the growth measurement, diagnoses may be made earlier, improving patient outcomes. As discussed in Chapter 1, there is considerable uncertainty in measurements by radiologists—given the same nodule, the radiologist will provide a slightly different measurement each time. In contrast, an automated method should consistently provide the same measurement. Many different automated methods have been developed, but there have been no studies which have compared automated methods to each other on the exact same dataset with identical evaluation methodologies.

The VOLumetric Change Analysis of NOdules (VOLCANO) study [55] was developed to compare different automated methods on the same dataset. The methods were compared for consistency in growth measurement. In developing the study, cases were selected to address the issues of varying slice thickness between scans, the use of zero-change nodules for evaluation, and the nodule characteristics that may affect measurement uncertainty. The evaluation methods for this study were designed to not rely on any human measurements, based on the understanding that human measurements are not reliable enough to serve as ground truth, and to analyze the growth measurement directly, instead of size measurements.

In this study, 18 algorithms from 13 research groups were evaluated on a database of 49 real nodules and one synthetic nodule. This study both bench-

marked the current state of the art in nodule growth measurement and provided the framework by which new algorithms may be evaluated.

2.1 VOLCANO study design

In the evaluation of automated nodule growth measurement algorithms, we are interested in the accuracy of their growth measurements in a wide variety of pulmonary nodules. This is made difficult by the lack of available ground truth, as was described in Section 1.4. We addressed this lack of ground truth by using two surrogate approaches: measuring the variability of the algorithms using nodules that have zero-change and synthetic nodules with a known change in size. Zero-change nodules allow for the characterization of bias, but it is trivial for an algorithm to produce a result of 0 and algorithm performance may vary from the zero-change case. Synthetic nodules provide a means to verify the accuracy of the algorithms, since the size change and density are known, but they fail to fully capture the complexity of a nodule *in vivo* – real nodules have attachments, density inhomogeneity, and varying intensity along their margins. Real nodules with growth were also included in the database; these nodules have pairs of scans at a sufficiently large time interval to ensure that actual growth occurred.

Taking these considerations into account, the study involved measuring the change in nodule size for 50 scan pairs divided into the categories above. To assess how algorithms are affected by a change in slice thickness, the zero-change nodules were divided into two groups—nodules imaged on scans of the same slice thickness and nodules imaged on scans of different slice thickness. Thus, the data may be divided into four subgroups:

1. (14) zero-change in which the scans were taken minutes apart and there-

fore there is no real change in the nodule size.

2. (13) zero-change cases as in A above except one scan had a slice thickness of 1.25 mm and the second scan had a larger slice thickness (2.5 or 5.0 mm)
3. (19) nodules with a significant time interval between scans and therefore some real change and (3) nodules with a large change in size of greater than 150%, one of which was known to be malignant. Of these nodules, 19 were considered to be stable or benign by biopsy and 3 were diagnosed as malignant.
4. (1) synthetic phantom nodule with a known size recorded multiple times with different slice thicknesses

Four additional scan pairs were made available to research groups for training. No information was provided regarding the subgroup of any of the nodules, but participating research groups in the study were aware of the existence of the four subgroups. The participants reported the fractional change in nodule size for each of the 50 scan pairs. Thirteen different participants submitted their measurement change results from on a total of 18 different methods. In 12 of these methods, the actual volumes recorded for each nodule were also reported.

In general, the main interest is to learn the smallest size change which can be reliably detected and the precision of that size change measurement. A number of studies on repeat scans have been reported in the literature [44, 45, 46, 47, 48], which are described in Section 1.4.4, and in these studies, the limits of agreement for repeat scans of the same nodule are on the order of 20-25% by volume. For these reasons, most of the cases in data set C were selected to have a size change within the range of $\pm 50\%$. For completeness, three cases with a very large size change (150% or more) were included to characterize the measurement methods for such situations.

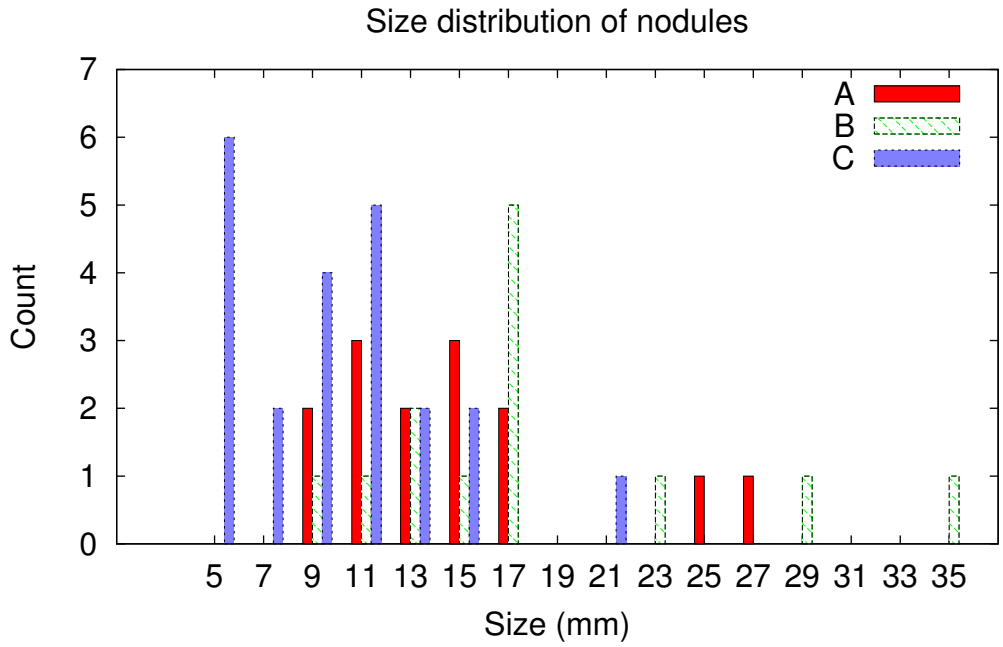


Figure 2.1: Size distribution of nodules in the evaluation dataset

2.1.1 Image data and preparation

The image data used in the study were acquired for the Public Lung Database to address drug response [56] and were provided by the Weill Cornell Medical College with the exception of one case of a synthetic "phantom" nodule provided by the FDA [38]. Cases were selected that contained at least one nodule of solid consistency which was present in at least two scans with a whole-lung field of view; only nodules visible on at least three slices on both scans were included. The size distribution for the real nodules used in this study is shown in Figure 2.1.

The scan pairs in group A had the same slice thickness for both scans; in 13 cases, both scans had a slice thickness of 1.25 mm while one case had scans with a slice thickness of 2.5 mm. For group B, 11 cases had scans with 1.25 mm and 2.5 mm slice thickness, while two cases had scans of 1.25 and 5.0 mm slice

Table 2.1: Summary of scan parameters for nodules in the dataset

Group	Current (mA)	Scanner Models
A	40-250	GE LightSpeed {Ultra, QX/i, Pro 16}
B	20-80	GE LightSpeed {Ultra, QX/i}
C	40-80	GE LightSpeed {Ultra, Pro 16, VCT}
D	40	Philips MX8000 IDT 16

thickness. All the scans in group C had a slice thickness of 1.25 mm. Scans in these three groups were acquired without overlap. All scans were acquired using a voltage of 120 kVp. The current and scanner model for each of the groups are listed in Table 2.1.

For group C, the status of each nodule was determined by a radiologist; stable nodules were either biopsied (3) or did not have any clinical change in 2 years (16), while the three malignant nodules were biopsied.

Group D was comprised of the synthetic phantom nodule, a 10 mm (523.60 mm^3) sphere with two different slice thickness reconstructions, 1.5 mm and 3.0 mm with 50% overlap. All scans have a whole-lung field of view. The phantom was placed in a chest phantom with simulated vascular structures [38].

In five zero-change cases, the patient position changed between prone and supine in the scans; one case was in group A while four cases were in group B.

Prior to making the images available for the study, all identifying patient information was removed. The original dates of the scans were replaced with dates corresponding to a time interval of 100 days between scans, with the order of the scans was randomized. The scans were then clipped in the axial direction, with the five slices below and above the region containing the nodule included if possible. This was done because some of the scans did not cover the whole-lung in the axial dimension and to reduce the amount of data to be downloaded

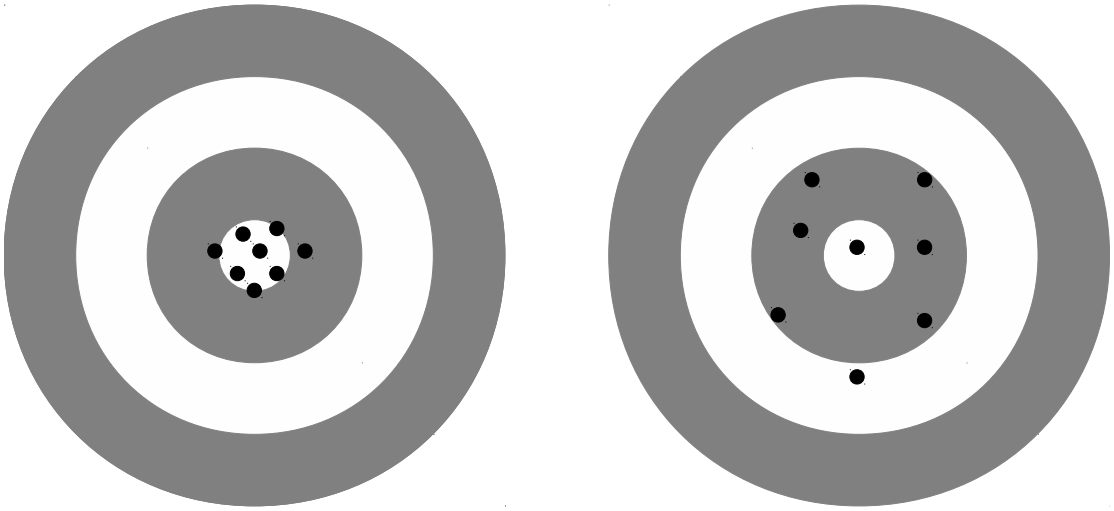
for this study. Along with the scans, participants were provided with a spreadsheet identifying the approximate center of the nodule established by a human observer.

2.2 VOLCANO evaluation metrics

We seek to quantify the **bias** and **variation** when evaluating a nodule growth measurement algorithm. In general, the bias measures the deviation from the true value, while the variation measures the spread of observations. The distinction between bias and variation is illustrated in Figure 2.2; the true value is indicated by the center of the "bulls-eye". In the context of nodule growth measurement, the bias of an algorithm is a measure of its deviation from the true growth of the nodule. Variation is the range or "dispersion" of growth measurements for nodules with identical growth. For example, given multiple scan pairs of a set of synthetic nodules with a known, fixed difference in size, bias would characterize the deviation of an algorithm from the true change in size, while variation characterizes the range of measurements. In the ideal case, both bias and variation would be zero.

To quantify the bias and variation of nodule growth measurement algorithms, the change in size is required. While the time span of the size change is also required to measure growth, we assume that any bias and variation will be due to the size measurement and not the time measurement. In the VOLCANO study, participants were requested to provide a size change metric for each nodule, from which metrics were derived to quantify the bias and variation.

For methods that measure the volume of the nodule on each scan, the size



(a) Low bias and low variation

(b) Low bias with high variation

(c) High bias with low variation

(d) High bias and high variation

Figure 2.2: Illustration of the distinction between bias and variation and the four possible combinations of low and high bias and variation

change metric is the relative volume change (RVC):

$$\text{RVC} = \frac{V_2}{V_1} - 1 \tag{2.1}$$

where V_2 and V_1 are the volumes measured for the nodule on scan 2 and scan 1 respectively.

The metric that was requested for results (relative change with respect to time 1) is asymmetric with respect to the order of presentation: no change in size has a value of 0, a 10 times increase in size results in a value of 9 while a 10 times decrease in size results in a value of -0.9. The negative size change is bounded by -1 while the positive size change is unbounded. In the dataset provided to the participants, the order of the scan pairs was randomized. It is possible to reverse the order, that is computing the relative size change from scan two to scan one by the following transformation:

$$\text{RVC'} = \frac{1}{RVC + 1} - 1 = \frac{V_1}{V_2} - 1 \tag{2.2}$$

where RVC represents the reported relative size change reported by the participants. However, if any of the methods has an order bias, adjusting the results to match the correct time sequence would mask such an effect. The one data set where ordering is important is data set B which had scans of different slice thicknesses; we transformed the results so that the thin slice scan was always the first in the pair to determine the effect caused by a change in slice interval. For the other datasets, transforming the RVC to restore the correct temporal order of the scans was not performed.

This study provides an opportunity to analyze several aspects of automated nodule change measurement performance. We can draw conclusions about the performance of the algorithms on the different nodule groups (A-D), compare the algorithms to each other, and attempt to determine what nodule character-

istics may affect the performance of the algorithms. The evaluation metrics are described in the following sections.

2.2.1 Statistical descriptors of general algorithm performance for each nodule group

One facet of automated nodule change measurement performance we seek to study is how the algorithms, in general, behave on datasets with different characteristics. In the case of zero-change nodules, where the scans are taken minutes apart from each other, the true growth is zero. Based on this, we can measure both the variation and bias of the methods on the zero-change datasets, group A and B. For the nodules in group C, the true size change is unknown, so it is difficult to interpret the bias, but it is provided for consistency. Finally, for group D, the phantom nodule, the true size and volume change is known.

Given the relatively small number of nodules in each category, non-parametric statistical descriptors were used in this study, since the assumption of normality may not be satisfied. To quantify the variation and bias of each group, the median of the median of absolute deviation (MMAD) and median of the absolute median (MAM) were computed for each group of nodules. The median of the absolute deviation (MAD) is computed by taking the median size change metric for each nodule (across all methods), and, for each method, computing the absolute deviation from the median. This results in a MAD value for each nodule. The median of the MAD values of the nodules in a group is the MMAD for the group. Given the set S_i of size change measurements for m methods for a nodule i ,

$$S_i = \{s_{i,1}, s_{i,2}, \ldots, s_{i,m}\}$$

the median size change (MS) is

$$\mathrm{MS}_i = \mathrm{median}(S_i)$$

the absolute deviation for each method a is

$$|\mathrm{D}_a| = |s_{i,a} - \mathrm{MS}_i|$$

and the median of the absolute deviation is

$$\mathrm{MAD}_i = \mathrm{median}(|\mathrm{D}_1|, |\mathrm{D}_2|, \ldots, |\mathrm{D}_m|)$$

Let there be n nodules in each group, then the MMAD for a particular group is

$$\mathrm{MMAD} = \mathrm{median}(\mathrm{MAD}_1, \mathrm{MAD}_2, \ldots, \mathrm{MAD}_n)$$

A higher MMAD value indicates a group of nodules that has higher variation.

To quantify the bias from zero, the MAM is computed by taking the median of the absolute median size change metric for each nodule:

$$\mathrm{MAM} = \mathrm{median}(|\mathrm{MSC}_1|, |\mathrm{MSC}_2|, \ldots, |\mathrm{MSC}_n|)$$

The MAM provides an estimate of the bias for a particular group. For zero-change nodules, the methods should report a size change metric of 0, which would result in a MAM of 0 as well. Nodules that have a bias will have positive MAM values. Note that these metrics are computed for each nodule and group; these are not descriptors of the performance of a specific algorithm.

2.2.2 Graphical comparison

The MAM and MMAD provide statistical descriptors of the different groups of nodules. To gain a better understanding of how the methods behaved on individual nodules, box-and-whisker plots were made for each nodule in all

the datasets. These plots provide a graphical representation of the following data:

- median (dark line in center of box)
- 25th and 75th percentiles (top and bottom of box)
- the highest and lowest values within 1.5 times the interquartile range (IQR) (top and bottom whiskers)
- outliers that fall outside 1.5 times the IQR (circles)

The median for a nodule would indicate the most likely measurement of the size change, and hence the expected bias. The size of the box is related to the amount of variation between the methods, with larger variations producing larger boxes. The whiskers and outliers also serve as a measure of variation, with whiskers that are farther out or a large number of outliers indicative of more variation. These plots can be used to analyze both the behavior of the algorithms on specific nodules (by comparing nodules to each other) and the performance of algorithms relative to each other.

2.2.3 Statistical comparison of different nodule groups

The statistical agreement between methods was established for the size change measurements using the Friedman test. The Friedman test is a non-parametric test that, in this study, assess whether or not the size change measurements from the different methods are statistically different across an entire group of nodules. To compare the behavior of groups, the Wilcox rank-sum test was applied to the MAD values computed for each nodule. This test assesses whether the biases are statistically different between the groups.

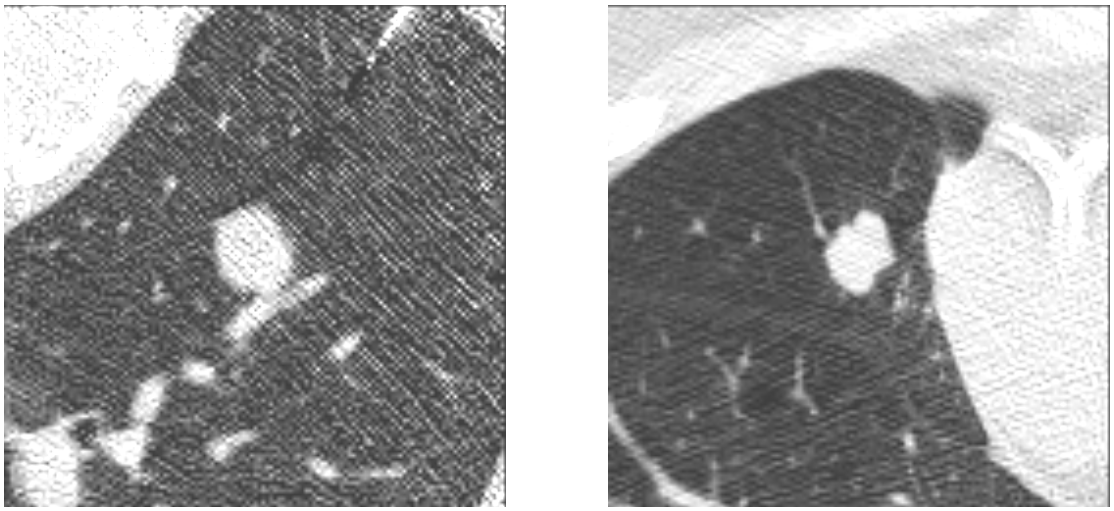
2.2.4 Comparison of methods

In addition to the analyses described above, which primarily seek to measure the performance of automated algorithms as a group, additional analysis can be performed to attempt to distinguish methods from each other. With the available data and lack of ground truth, the methods were compared to each other based on the number of times their measurements were outliers compared to the entire group. However, this is not a measure of whether the algorithm is providing the correct measurement; it serves only to distinguish methods from each other.

2.2.5 The impact of nodule characteristics

We hypothesize that nodule characteristics can be identified that affect the measurement performance of an automated nodule growth measurement algorithm. A previous study by Wang et al. (2008) [49] found that the variation in size measurement of nodules was affected by nodule location, morphology, and size. Since this study includes many different algorithms, we can assess whether there are nodule characteristics that universally affect automated algorithms. To address this, the presence of various nodule characteristics was determined by visual inspection of the nodules.

The examined characteristics can be divided into two categories: 1) characteristics that are attributable to the scanner and 2) nodule morphology. There are four characteristics affected by the scanner: slice difference (SD), which for the nodules in group B was the difference in the slice thickness for the nodules in group B, or the difference in the number of slices on which the nodule appeared for group A; whether the X-ray tube current was altered between scans (CC); the presence of visible noise (N); and whether the position of the patient

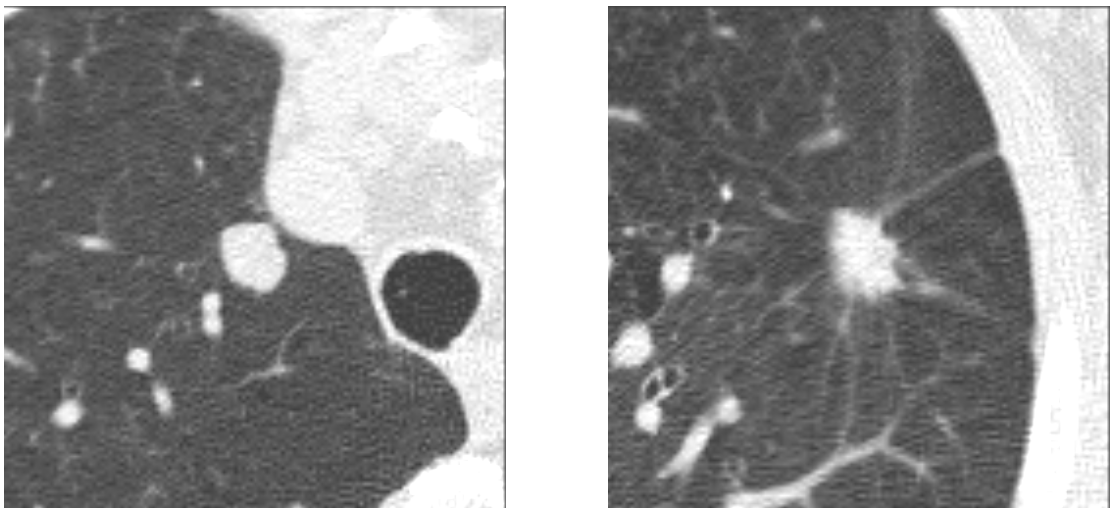


(a) (b)

Figure 2.3: Examples of a nodule a) with visible noise in the scan and a blood vessel attachment and b) without noise or attachments.

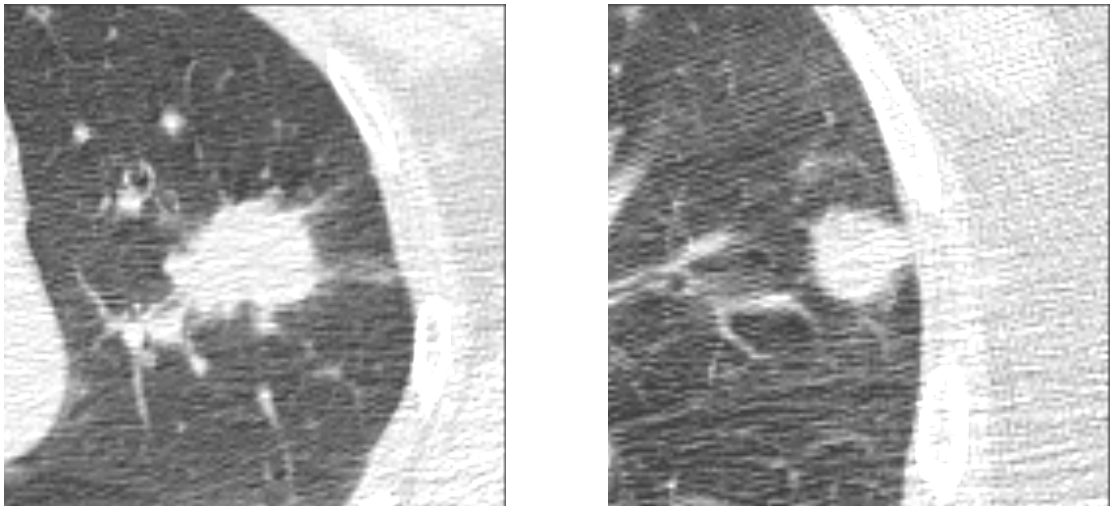
was altered between scans (PC). The presence of visible noise was based on visual observation and judgment by the author; a comparison between a nodule with noise and without noise is shown in Figure 2.3.

The three nodule morphology characteristics examined were the margin sharpness (M), the presence of attachments (A), and the presence of spiculations (S). The margin was either ill-defined (I) or well-defined (W), assessed by visual observation of the author. Nodules with well-defined margins have a sharp demarcation between the nodule and lung parenchyma, while nodules with ill-defined margins have a larger zone of transition between the nodule and lung parenchyma that may increase the uncertainty in growth measurement. A comparison of the two nodule types is shown in Figure 2.4. The nodules were observed for attachments to either blood vessels or the chest wall, an example of which is shown in Figure 2.3a, as well as spiculations, which are irregular points or spikes on the nodule surface, shown in Figure 2.5.



(a) (b)

Figure 2.4: Examples of a nodule with a) well-defined margins and b) ill-defined margins. The margin is the size of the transition region from the nodule to the lung parenchyma, and nodules with ill-defined margins will appear to have a more blurred boundary.



(a) (b)

Figure 2.5: Examples of a nodule a) with spiculations and b) without spiculations.

For each nodule, the bias and variation were computed across methods. The bias for a nodule i was computed as the median of the size change measurements provided by all the methods for each nodule, MS_i. The variation was computed as the median of the absolute deviation, MAD_i. For the nodules in group C with real growth, the bias was not included because the true change in size of the nodule was unknown. The nodules were ranked according to their bias and variation within their groups. The Wilcoxon rank sum test was used to assess the statistical significance of each characteristic on both the bias and variation.

2.3 Automated methods included in the study

Thirteen research groups participated in the VOLCANO challenge. Several groups submitted multiple size change measurement methods for a total of 18 submitted methods. Although not required, 12 groups provided volume measurements for each nodule. The research groups and their methods are summarized in Table 2.2; methods provided by the same group have the same prefix. The category of methods are described in additional detail below.

There were a wide variety of methods in the VOLCANO challenge. To make comparison between the methods easier, we ranked the methods based on the amount of operator interaction required and the types of algorithms employed by the method.

The levels of operator interaction were divided broadly into three groups: completely automated after specification of a seed point, manual parameter control, and modification of the resulting boundary or indicating additional control points. These categories were further subdivided according to the frac-

Table 2.2: Summary of participating methods (PVC = partial voxel correction)

	Method	Automation	Category
1	Tokushima	3	Image filtering
2	ISI-Sphere	4	Sphere fitting
3	ISI-Seg	4	Image filtering
4	ISI-Reg	4	Elastic registration
5	NYU-HYB	7	Image filtering with PVC
6	NYU-HYBA	7	Image filtering with PVC
7	UCLA	6	Image filtering
8	VIA-GAD	1	Density change
9	VIA-GAS	4	Image filtering
10	Kitware	4	Fast marching and shape detection level set
11	Duke	1	Spiral scanning, dynamic programming
12	Gifu	1	Image filtering
13	Biomedsys	2	Image filtering
14	MeVIS	3	Image filtering with PVC
15	Siemens	3	Image filtering with PVC
16	Philips	1	Active contour
17	Definiens	5	Image filtering
18	VIA-ZCOMP	4	Image filtering with Z-compensation

tion of cases which required manual intervention. Groups were asked to rank the level of automation required by their algorithms using the following scale:

1. Totally automatic using seed points
2. Limited parameter adjustment (on less than 15% of the cases)
3. Moderate parameter adjustment (on less than 50% of the cases)
4. Extensive parameter adjustment (more than 50% of the cases)
5. Limited image/boundary modification (on less than 15% of the cases)
6. Moderate image/boundary modification (on less than 50% of the cases)
7. Extensive image/boundary modification (more than 50% of the cases)

The scale is roughly in order of operator effort required, with methods requiring only a seed point using the least about of operator effort. The level of automation required for each method is detailed in Table 2.2.

In addition to the level of automation, the methods were also categorized according to the approach that was utilized. While the details of the methods all differ, we divided them into three main categories: image filtering, image filtering with partial voxel correction, and other approaches. These will be described in the following sections.

2.3.1 Image filtering approaches

Methods that used the image filtering approach for measuring nodule size change segmented the nodule using image filtering operations, such as thresholding, morphological filtering, and connected component analysis, and used the change in the segmented volume as the nodule size change. This approach was used by

seven methods (Tokushima, ISI-Seg, UCLA, VIA-GAS (Vision and Image Analysis Group, Cornell), Gifu, Biomedsys, Definiens) is based on image filtering operations [22, 23]. Generally, methods using this approach extracted a volume of interest (VOI) around the seed point for each nodule. This VOI was resampled and a threshold applied to identify voxels belonging to high-intensity structures. Next, either region growing or connected component analysis was applied to the volume of interest to eliminate non-nodule structures, followed by the removal of attached structures such as vessels or the chest wall using morphological filtering or other more advanced techniques. The volume for each nodule was computed from the voxel size and number of voxels included in the segmentation; the size change metric for these methods was based on the volume change. Additional details of the algorithms are provided below.

2.3.1.1 University of Tokushima (Tokushima)

The Tokushima approach to nodule segmentation relied upon several image filtering operations to identify the VOI and segment the nodule. The lung region is segmented based on thresholding and connected component analysis. Unique to this algorithm is a step that corrects for the background bias in the lung region, which is responsible for a gradually increasing density towards the periphery of the lung region. After the VOI is identified and background bias correction performed, the nodule is segmented by removing noise from the image, applying a threshold to the image, and removing attachments by morphological operations.

2.3.1.2 Image Sciences Institute Segmentation (ISI-Seg)

The ISI segmentation algorithm is very similar to the method described in Section 1.6, with the addition of a 2D lung segmentation prior to segmenting the nodule [57]. The 2D lung segmentation was performed using thresholding, connected component labeling, and morphological operations [58].

2.3.1.3 University of California, Los Angeles (UCLA)

The algorithm used by UCLA is a seeded-region growing technique. The user provides a seed point and a line from the seed point to the lung parenchyma. Voxel intensities along the line are sampled to form a histogram from which a threshold is determined to separate the nodule from the background. Using this threshold, a 3D seeded region growing is performed starting from the user-specified seed point.

2.3.1.4 Cornell volumetric method (VIA-GAS)

The VIA-GAS method is based on thresholding followed by morphological filtering and vessel removal. It is described in detail in Section 1.6.

2.3.1.5 Cornell z-compensation method (VIA-ZCOMP)

The VIA-ZCOMP method uses moment analysis to compensate for variation in the z-dimension of the scan. This computes the moments on the segmented image from the VIA-GAS method, and is described in further detail in Chapter 3.

2.3.1.6 Gifu University (Gifu)

The method provided by Gifu University is a nodule segmentation method primarily based upon the refinement of a thresholded region and described in further detail by Hayashi, Zhou, and Fujita (2009) [59]. In the first stage of the algorithm, the lung region is segmented using thresholding, connected component analysis, and morphological filtering. The nodule is segmented by applying rules to the distance transform image to isolate vessel-like structures, followed by thresholding, connected component analysis, and morphological filtering.

2.3.1.7 Biomedical Systems (Biomedsys)

Biomedical Systems provided results from their LifeRx^{TM}volume tool. Using the tool, a user creates initial boundaries around the VOI; these boundaries need not be done on every slice. The user also provide seed points within the nodule, and a segmentation is performed based on the histogram of the VOI.

2.3.1.8 Definiens

The Definiens method is a 3D region growing-based algorithm implemented in the Cognition Network Language [60]. The lungs and other anatomical objects are first segmented from the image. Given a seed point, a seed object is segmented from the image based on similar image intensities and proximity. Based on this seed object, the histogram is analyzed to determine the lower and upper bounds for the intensity. The seed object is grown using the intensity information and adaptive surface tension. Finally, morphological filtering is performed to refine the segmentation. The algorithm was developed primarily for large tumors and trained with cases with a variety of resolutions. Although only a seed point needs to be specified, manual editing of the boundary is possible.

2.3.2 Image filtering with partial voxel correction

Four methods (MeVIS, NYU-HYB, NYU-HYBA, Siemens) extended the image filtering approach to better address partial voxels along the perimeter of the nodule [24, 25]. These methods determined a region around the border of the nodule where the voxels have intensity between that of solid tissue and the lung parenchyma; based on histogram analysis, these voxels were weighted when computing the nodule volume.

2.3.2.1 New York University (NYU-HYB and NYU-HYBA)

The two methods used by NYU both performed additional analyses on the voxels at the nodule-lung parenchyma interface [24]. First, the methods begin by generating a small seed region at the center of the nodule. Voxels are selected by applying a threshold determined from this small seed region. Morphological erosion is performed on the region, followed by connected component analysis and region growing. The growing operation is devised to over-segment the nodule and include some of the surrounding structures, designated as region I. Next, an adaptive threshold is determined by finding the midpoint of the intensities of voxels on the edge of the region and voxels towards the center, resulting in a region J.

In the NYU-HYB method, region J is eroded by 3 voxels, forming a new region H. The region between I and H, that is, $I - H$, is a hollow shell that should contain voxels along the boundary between the nodule and lung parenchyma; some of these voxels are likely to be partial voxels. A portion of the volume of these voxels can be included in the nodule volume based on their intensity relative to the intensities of the region.

The NYU-HYBA is a variant of the NYU-HYB method that uses fixed inten-

sities thresholds instead of determining the values from the voxel intensities of the region.

2.3.2.2 MeVIS

The method presented by MeVIS is described in detail by Kuhnigk et al. (2006) [25]. Much like the NYU methods, additional steps are performed to achieve better accuracy along the boundaries of the nodules; after segmentation, the region along the boundary of the nodule is identified and voxels within this region contribute partially to the volume of the nodule.

2.3.2.3 Siemens

Siemens provided results from an algorithm that is available as part of one of their commercial products, Oncology, Syngo MMWP VE31 A. The algorithm is initialized using either a seed point or by providing a line across the largest diameter of the nodule. A VOI is extracted from the CT scan, and 3D region growing is performed starting from the seed point or line. Morphological operations are used to separate the nodule from any attached structure. Finally, a check is performed of the resulting segmentation to ensure that the location of the initial conditions is sensible, relative to the segmentation. If not, the algorithm repeats with slightly different thresholds.

2.3.3 Other approaches

The remaining six methods used different approaches. Most methods resampled the CT scans into isotropic space.

2.3.3.1 ISI-Sphere

ISI-Sphere estimated the best fitting spherical volume of interest at the seed point of the nodule from a thresholded, resampled volume of interest [57]. The volume of the nodule was estimated from the number of voxels above a threshold.

2.3.3.2 ISI-Registration

The ISI-Registration method applied non-rigid registration to deform the first scan to the second; this transformation was then applied to a segmentation obtained for the first scan to obtain the volume of the nodule on the second scan [57].

2.3.3.3 Kitware

Kitware required only a manually specified seed point and a bounding box. Their method computed several features for each voxel, including lung wall, vesselness, gradient, and intensity features which were aggregated and used to guide a fast marching algorithm to generate an initial guess of the nodule boundary. The initial boundary estimate was refined using a shape detection level set. The volume was computed from the surface of the resulting level set.

2.3.3.4 Duke

Duke used a spiral-scanning technique to convert the 3D volume of interest around the nodule to a 2D generalized polar coordinate system. Dynamic programming techniques were used to obtain the nodule boundary on the 2D image which was then transformed back into 3D space [27]. This boundary was applied to the original 3D image to estimate the nodule volume.

2.3.3.5 Philips

Philips used an active contour-based approach utilizing a radial basis function energy minimization algorithm [26].

2.3.3.6 VIA-GAD

In contrast to the methods described thus far, VIA-GAD (Cornell) did not explicitly segment the nodule; instead, the change in nodule size was estimated from the change in density of a Gaussian-weighted region around the nodule [61]. This method is described in detail in Chapter 4 as the density growth index method using a Gaussian weight (DGI_G).

2.4 Results

The results of the study can be divided into three broad categories: group results, individual method results, and nodule results. Group results indicate the behavior of all the automated methods for each group of nodules, individual method results showcase the performance of each method, while nodule results provide information on the behavior of the methods for each nodule.

2.4.1 Group results

Box and whisker plots, plotted for the size change measurements for the different groups, are shown in Figure 2.6 to 2.10a. Note that in these plots, the x-axis represents a random identifier that differs from the case ID provided to the participants. For group C, due to the large change in three nodules, two plots are

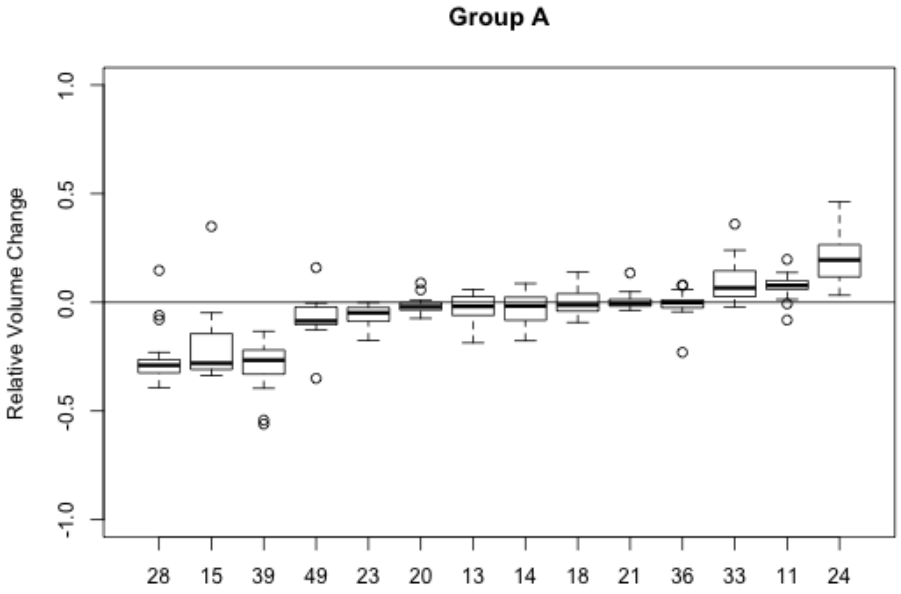


Figure 2.6: Box plot for group A: zero-change, same slice-thickness. For each nodule, the median of the relative volume change (RVC) is plotted as the dark solid line within the box. The lower and upper bounds of the box represent the 25th and 75th percentile of RVC, and the whiskers indicate the lowest and highest RVC within 1.5 times the interquartile range. Any RVC measurements that lie outside 1.5 times the interquartile range are indicated by circles.

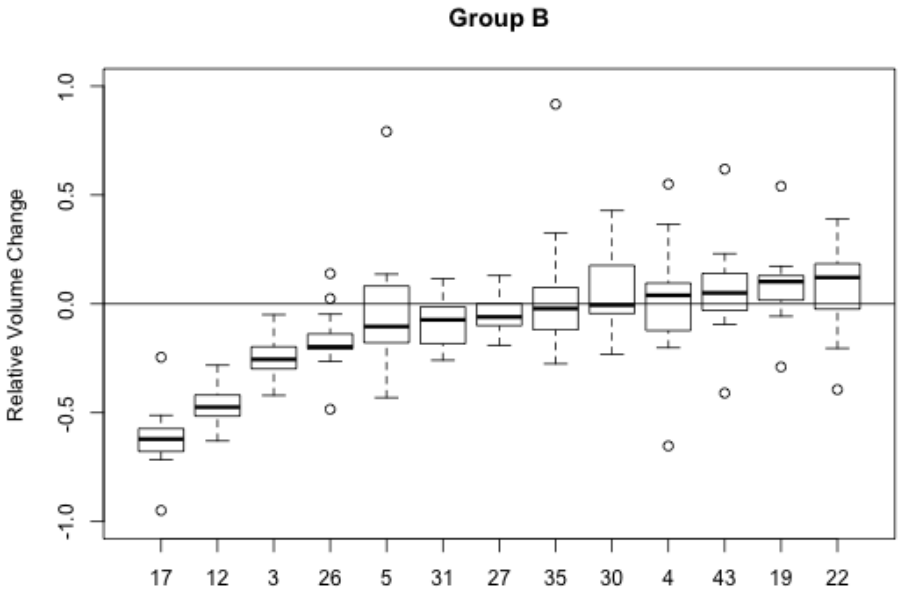


Figure 2.7: Box plot for group B: zero-change, different slice thickness. The volume change was computed so that the first scan always had a slice thickness of 1.25 mm, while the second was 2.5 mm except for cases 3 and 21 which was 5.0 mm.

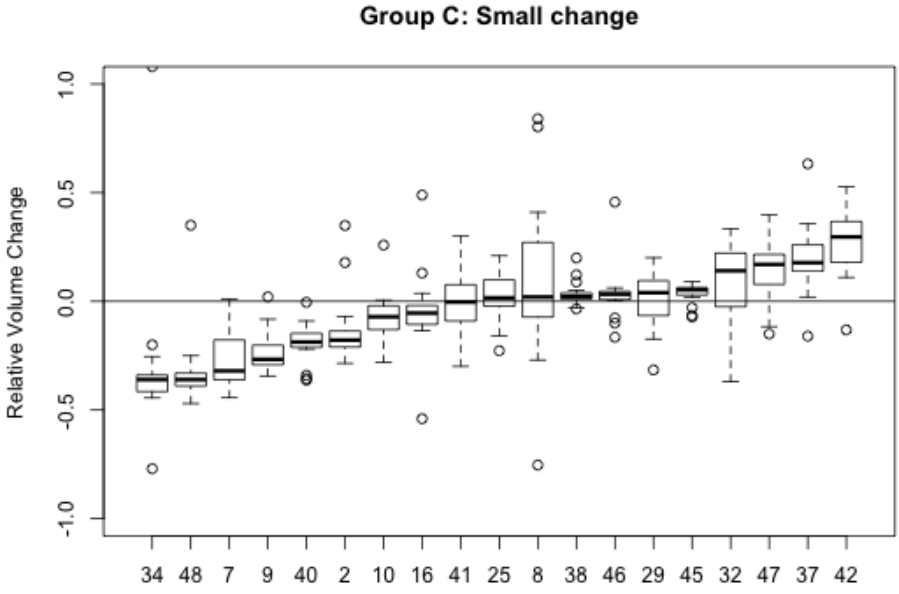


Figure 2.8: Box plot for group C: actual small change (one outlier not shown)

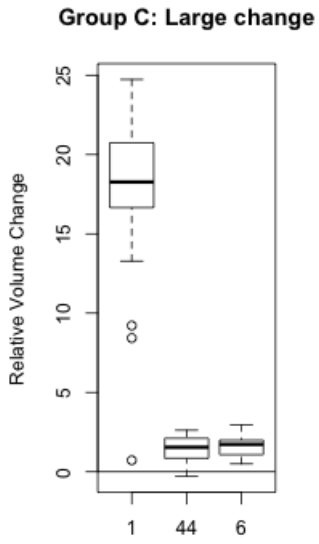
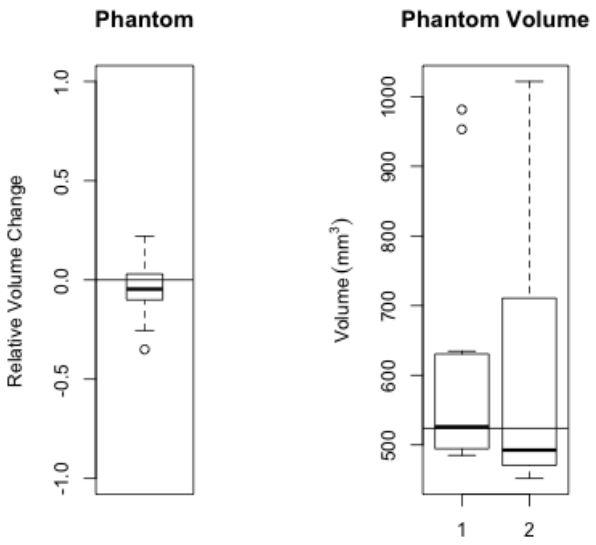


Figure 2.9: Box plot for group C: actual large change (the size change of the first nodule was inverted for visibility)

shown for clarity (Figures 2.8 and 2.9).

Although the primary focus of this study was on size change measurement, not volume measurement, participants were requested to provide volume measurements if available. Of the 18 methods submitted, 12 methods provided volume measurements. For the phantom, the only case where the true volume was known, a box plot is provided in Figure 2.10b for each scan. In the figure, the second scan was reconstructed at twice the slice thickness of the first scan.

Bias and variation results in the form of the median of absolute medians (MAM) and median of absolute deviation (MMAD) are provided in Table 2.3.



(a) Size change variation

(b) Volume estimates for 12 of 18 methods that provided size estimates. The horizontal line indicates the true volume of the phantom (523.60 mm^3)

Figure 2.10: Box plot for group D: phantom nodule

Table 2.3: Summary of the median of absolute median (MAM) and median of median of absolute deviation (MMAD) of the relative size change measurements for each group, which measure the bias and variation respectively. Note that the results of group D are on a single nodule.

	Group A	Group B	Group C	Group D
MAM	0.0554	0.0953	0.1719	0.0620
MMAD	0.0408	0.0842	0.0680	0.0490

Table 2.4: Number of outliers produced by each method, according to group

Method	Automation	Type	A	B	C	Total
Biomedsys	2	IF	0	0	3	3
Definiens	5	IF	0	3	6	9
Duke	1	O	0	0	1	1
Gifu	1	IF	0	0	0	0
ISI-Reg	4	O	5	3	5	13
ISI-Seg	4	IF	2	0	0	2
ISI-Sph	4	O	2	4	5	11
Kitware	4	O	3	2	2	7
MeVIS	3	IFP	1	0	0	1
NYU-HYB	7	IFP	1	0	1	2
NYU-HYBA	7	IFP	0	0	0	0
Philips	1	O	0	0	1	1
Siemens	3	IFP	1	0	0	1
Tokushima	3	IF	2	0	1	3
UCLA	6	IF	0	0	1	1
VIA-GAD	1	O	2	2	8	12
VIA-GAS	4	IF	0	0	1	1
VIA-ZCOMP	4	IF	0	2	2	4

2.4.2 Individual results

The size change measurements were reviewed to determine how many outlier measurements each method produced. An outlier was defined to be outside 1.5 times the interquartile range for the nodule. The results are shown in Table 2.4. This test only serves to show, generally, which methods showed more agreement with the other methods; it does not suggest that a method is closer to "truth".

2.4.3 Nodule and scan characteristics

The bias and variation, provided by the median and median absolute deviation (MAD), are given for each group in Tables 2.6, 2.8, and 2.10. The statistical significance of each of the nodule and scan characteristics are provided for each group in Tables 2.7, 2.9, and 2.11. The key for the column title abbreviations is provided in Table 2.5. The meaning of slice difference (SD) differs in group A and group B. In group A, SD indicates the difference in the number of slices on which a nodule appears on each scan. For example, if on one scan, the nodule appears on three slices, but appears on five slices on the second scan, SD would be two. Since the nodules in group A are supposed to represent zero change, SD should ideally be 0. In group B, since the slice thicknesses varied between the two scans, SD indicates the slice thickness of the two scans. SD is not relevant to the nodules in group C.

The volume of each nodule, as measured by the VIA-GAS algorithm, was also included in the table of nodule properties for each group. Neither the bias nor the variation was correlated with the nodule volume in group A (Spearman's rank correlation for bias: $\rho = 0.288$, $p = 0.318$; and variation: $\rho = -0.315$, $p = 0.273$) or in group B ((Spearman's rank correlation for bias: $\rho = -0.154$, $p = 0.617$; and variation: $\rho = 0.011$, $p = 0.978$). The variation in group C was also not significantly correlated with the volume, with a Spearman's rank correlation $\rho = -0.367$ and $p = 0.123$.

Table 2.5: Key for column headings; see Section 2.2.5 for more details.

Abbreviation	Meaning
Vol.1	Volume on first scan
SD	Slice difference
CC	Current change
A	Presence of attachments
N	Presence of visible noise
M	Margin: I=ill-defined, W=well-defined, M=Mixed
S	Presence of spiculations
PC	Position change of patient

Table 2.6: Bias and variation for group A and associated nodule properties

ID	**Bias (rank)**	**Var. (rank)**	**Vol.1 (mm^3)**	**SD**	**CC**	**A**	**N**	**M**	**S**	**PC**
28	-28.9% (1)	3.5% (8)	739.4	1	N	N	N	I	N	N
15	-28.0% (2)	5.6% (3)	265.9	1	Y	N	Y	W	N	N
39	-26.6% (3)	6.4% (2)	741.3	3	N	Y	N	I	Y	N
24	-19.5% (4)	7.8% (1)	620.1	0	Y	Y	N	W	N	N
49	-8.6% (5)	2.4% (10)	2859.3	0	N	Y	N	W	N	N
11	7.8% (6)	1.8% (14)	1626.7	0	N	Y	N	W	N	N
33	6.6% (7)	4.9% (5)	373.4	0	N	Y	Y	W	N	N
23	-4.8% (8)	2.9% (9)	1963.1	1	N	Y	N	W	N	N
20	-2.1% (9)	1.8% (13)	943.7	0	N	Y	N	W	N	N
13	-1.8% (10)	4.5% (6)	10146.1	0	N	Y	N	I	Y	N
14	-1.7% (11)	5.2% (4)	6966.1	0	N	Y	N	I	Y	Y
18	-1.2% (12)	4.2% (7)	2083.5	0	N	Y	Y	W	N	N
21	-0.7% (13)	2.1% (12)	1263.1	0	N	N	N	W	N	N
36	0.0% (14)	2.2% (11)	779.1	0	N	N	N	W	N	N

Table 2.7: Statistical significance of nodule properties for group A (Wilcoxon rank sum test)

	Bias	**Variation**
SD	0.008	0.288
CC	0.132	0.055
A	0.635	0.620
N	0.769	0.275
M	0.240	0.179
S	0.770	0.119
PC	0.857	0.456

Table 2.8: Bias and variation for group B and associated nodule properties

ID	**Bias (rank)**	**Var. (rank)**	**Vol.1 (mm^3)**	**SD**	**CC**	**A**	**N**	**M**	**S**	**PC**
17	-62.1% (1)	5.7% (9)	1629.9	1.25,2.5	N	Y	N	I	Y	N
12	-47.5% (2)	5.6% (10)	4027.2	1.25,2.5	Y	Y	Y	I	N	N
3	-25.5% (3)	5.8% (8)	2119.8	1.25,2.5	Y	Y	Y	W	N	N
26	-19.8% (4)	4.2% (13)	1114.0	1.25,2.5	Y	Y	N	W	N	N
22	12.1% (5)	10.6% (3)	334.0	1.25,2.5	N	N	N	I	N	N
5	-10.4% (6)	12.8% (2)	1288.1	1.25,2.5	Y	N	N	I	Y	Y
19	10.2% (7)	4.8% (11)	1346.0	1.25,2.5	N	Y	N	I	Y	N
31	-7.3% (8)	10.2% (5)	24194.7	1.25,2.5	Y	Y	N	W	N	N
27	-6.0% (9)	4.3% (12)	12918.9	1.25,2.5	Y	Y	Y	W	N	Y
43	5.0% (10)	9.0% (7)	2112.0	1.25,5.0	Y	Y	N	W	N	N
4	3.9% (11)	12.9% (1)	2545.2	1.25,5.0	N	Y	N	I	Y	N
35	-2.2% (12)	9.8% (6)	2095.1	1.25,2.5	Y	Y	Y	I	N	N
30	-0.6% (13)	10.5% (4)	5192.5	1.25,2.5	Y	Y	N	W	Y	Y

Table 2.9: Statistical significance of nodule properties for group B (Wilcoxon rank sum test)

	Bias	**Variation**
SD	0.231	0.308
CC	0.260	0.604
A	0.513	0.103
N	0.260	0.260
M	0.836	0.366
S	0.833	0.284
PC	1.000	0.692

Table 2.10: Bias and variation for group C (small change) and associated nodule properties

ID	**Variation (rank)**	**Vol.1 (mm^3)**	**CC**	**A**	**N**	**M**	**S**	**PC**
8	23.5% (1)	48.5	N	Y	N	W	N	N
29	10.5% (2)	28.4	N	N	N	W	N	N
32	10.2% (3)	39.5	N	Y	N	W	N	N
42	9.0% (4)	326.8	N	Y	N	I	Y	N
41	8.3% (5)	485.4	N	Y	N	I	Y	N
37	8.2% (6)	1089.7	N	Y	N	I	N	N
25	8.1% (7)	573.6	N	Y	N	W	N	N
7	7.2% (8)	107.4	N	N	N	W	N	N
47	5.9% (9)	854.9	N	N	N	I	Y	N
10	5.9% (10)	731.3	N	N	N	I	N	N
16	5.2% (11)	188.9	Y	N	N	W	N	N
9	5.0% (12)	7025.3	N	Y	N	W	N	N
2	3.8% (13)	1841.1	N	Y	N	W	N	N
40	3.7% (14)	573.7	N	Y	N	W	N	N
34	3.7% (15)	101.8	N	N	N	W	N	N
48	3.0% (16)	356.7	N	N	N	W	N	N
46	2.1% (17)	66.7	N	N	N	W	N	N
38	1.6% (18)	1258.1	Y	N	N	W	N	N
45	1.5% (19)	484.3	N	N	N	W	N	N

Table 2.11: Statistical significance of nodule properties for group C (Wilcoxon rank sum test). Properties that were the same for all nodules in the group are indicated with a "-".

	Variation
CC	0.259
A	0.050
N	-
M	0.151
S	0.198
PC	-

2.5 Discussion

This study was the first to compare different automated nodule size change measurement methods on the same set of pulmonary nodules. Thanks to participation by several research groups, a total of eighteen different methods were represented in this study. From these measurements, we can make some observations about the performance of automated methods in general and determine some of the factors affecting size measurement.

2.5.1 Repeat measurement behavior

Zero-change datasets have been used in many previously published studies to quantify nodule size measurement variability. This dataset was represented by the nodules in group A. The variation between the methods, measured by the median MAD of 0.0408, indicated low variation between the methods. The 85% confidence interval of the absolute medians across methods (i.e., omitting the two largest cases) was 0.266 or a 26% volume change, which compares favorably to the variation reported by previous published studies of approximately 20%.

The bias from the true zero value of size change provides some insight into the accuracy of the size change measurement. For group A, the bias, as repre-

sented by the median absolute median, was 0.0554, which is a large deviation from zero. The non-zero bias is also apparent in the box plot for group A in Figure 2.6; for 8 of the 14 cases, the interquartile range (IQR) did not include zero. Therefore, there is evidence of a systematic bias introduced by the scanner/nodule combination for these cases.

To examine the variation due to the nodule and/or scanner, the median MAD was computed across nodules instead of across methods. The value of 0.0700 was larger than the median MAD computed across methods; thus, the variation due to the nodule/scanner is larger than the variation due to the methods. In addition, the median of the absolute medians, which measures the bias, was lower, 0.0202, suggesting that any bias that exists is due to the nodule/scanner and not the method.

2.5.2 Impact of change in CT slice thickness

One parameter that sometimes varies between scans when making a size change measurement is the slice thickness. Comparing the results of the methods on the nodules in group A, where the scans had the same slice thickness, and group B where there was a change in slice thickness, provides an indication of how the size change measurement is affected in such a scenario. The size change measurements across the different methods in group A for all the methods were in agreement ($p = 0.81$ using the Friedman test), that is, no method was significantly different from any other method. In contrast, the size change measurements for group B, where there is a change in slice thickness, were significantly different between the methods ($p < 0.01$).

To determine if the variation is greater in group B, the median absolute deviation (MAD) was computed for each nodule in group A and group B and the

Wilcoxon rank-sum test was performed. The variation was higher in group B than group A ($p < 0.01$); with $p = 0.05$, the variation is 35% higher in group B than group A. We also observe from Figure 2.10b that there is a greater dispersion in the IQR of the volume measurement for the thicker slice scan of the phantom nodule (131.01 mm3 vs. 207.79 mm3) which is consistent with the above results.

2.5.3 Variation of methods in presence of change

While the zero-change datasets enabled measurement of variation and bias of the methods, the results may differ when nodules with actually change are considered. To determine if the behavior of the methods was similar, the variation in measurements on the group C nodules (which had actual change) were compared to the variation in group A. As before, the variation was measured by the median of the median of absolute deviation (MMAD). The MMAD values for group A and the small change cases in group C were 0.0408 and 0.0655 respectively. These numbers are fairly similar and suggest that the results obtained from a zero-change dataset capture similar behavior of a dataset with a small amount of change.

There were three nodules in group C that exhibited a large change in size. For these nodules, there was an increased variation between the methods (MMAD of 0.4446). Although the number of cases was too small to perform statistical analysis, with larger size changes, there is a greater possibility of disagreement between the methods.

2.5.4 Comparison of volume estimation

In this study, the measurement of interest was size change, and not the actual size of the nodule, but the participants were requested to provide volume measurements to enable a comparison. For the 12 methods that provided volume information, the variation in the size change measurement was compared to the volume measurements for group A and the small change subset of group C. The Friedman test was performed to determine if the methods were in agreement. In group A, there was no significant difference between the size change measurements of the methods ($p = 0.92$), but there was significant disagreement between the volume measurements of the methods ($p < 0.01$). For the small change subset of group C, again there was no significant difference between the size change measurements of the methods ($p = 0.97$)), but significant disagreement in the volume measurements ($p < 0.01$). These results indicate that conclusions drawn from size measurements may not apply to size change measurements; one possible explanation is that a bias in size measurement between methods might be neutralized when computing size change.

2.5.5 Impact of automation level and algorithm type

This study included algorithms with varying approaches to nodule growth rate measurement, and this allows for observations about the general approaches, such as how does the level of automation affect the performance, and is one approach significantly better than the others? Methods which require less human intervention may not be able to correctly measure all nodules, leading to catastrophic errors in measurement, but for those nodules that are correctly measured, the performance should be better. Different approaches to growth measurement may have different tradeoffs that only become apparent in compari-

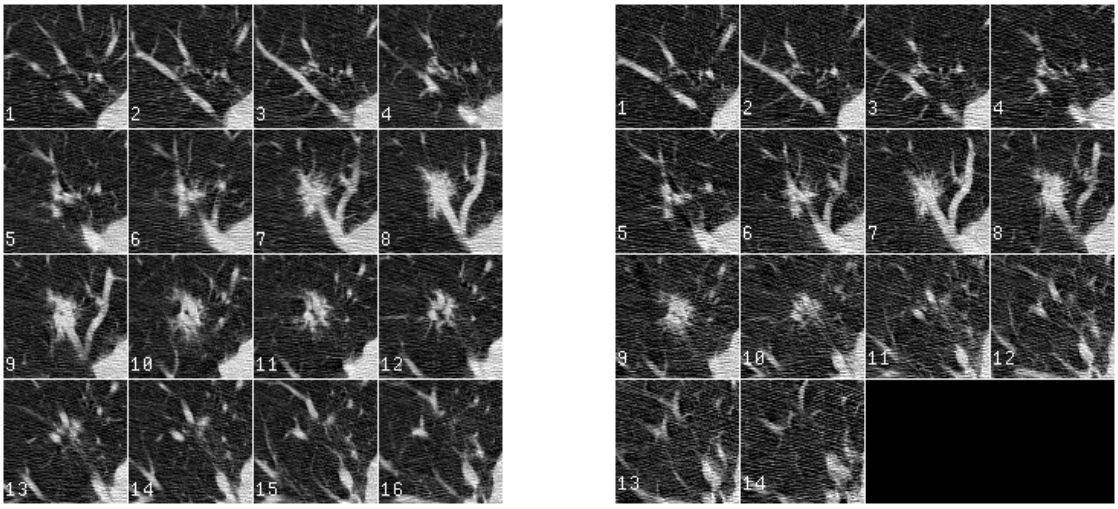
son to other methods. The analysis of the methods was based on their number of outlier measurements compared to the other methods.

There was no clear correlation between the level of automation of a given method and how it compared with the other methods. The two methods with zero outliers were the Gifu and NYU-HYBA methods, which represented the two extremes of automation. While the Gifu method required only a seed point, the NYU-HYBA method required boundary modification on a majority of cases. There were six methods with one outlier each, and the level of automation for these methods ranged from one to six. Thus, based on these 18 methods, the automation level of a method did not correlate with its performance in this study.

There were several different categories of algorithms, but image filtering with partial voxel correction (IFP) methods seemed, as a group, to have the fewest number of outliers; the highest number of outliers for an IFP method was 2, compared with 9 for image filtering (IF) methods and 13 for other methods. However, there were IF methods and other methods which had low numbers of outliers, so we can not say that one specific approach is better than another.

2.5.6 Scan and nodule characteristics that influence bias and variation

Determining the impact of nodule characteristics on the measurement of size change has implications for interpreting the results of automated methods and might be used to determine a level of confidence in the measurement. One previous study by Wang et al. (2008) [49] found that nodule morphology, location, and size all influenced the measurement variability of an automated method. In this study, we considered four possible sources of variation: scanner param-



(a) (b)

Figure 2.11: Example of a nodule from group A (zero-change, same slice thickness) with a three slice difference between the first scan a) and the second scan b)

eters, margin sharpness, presence of spiculations, and the presence of vessel attachments.

In group A, the three nodules with the greatest bias all appeared on a different number of slices on the two scans, with as much as a three slice difference between scans. The significance of this observed effect from the difference in slices was confirmed with $p = 0.008$. Two of the three nodules with the greatest variation had a difference in the current used in the two scans; these two nodules were also among the four nodules with the greatest bias. The change in current was almost statistically significant, $p = 0.055$. None of the other characteristics had statistically significant effects on either the bias or variation. Although the characteristics were only tested individually for statistical significance, spiculations in conjunction with a ill-defined margin tended to increase the variation, with the second, fourth, and sixth nodules with the highest variation having both characteristics.

It was difficult to find any clear trends in the data for group B, with no characteristics having any statistically significant effect on either the bias or variation. The difference in slice thickness seemed to dominate over all other characteristics, although there was no statistical difference between the nodules with 1.25 mm and 2.5 mm scans and those with 1.25 mm and 5.0 mm scans.

In group C, since the true size change was unknown, the bias could not be computed. The presence of attachments was the only significant effect ($p = 0.05$) on the variation. This is most likely because the algorithms all have different approaches to segmenting sections of the nodule that are attached to other structures; the presence of attachments was not significant in groups A or B due to the lack of nodules without attachments in those groups.

2.5.7 Future extensions for VOLCANO

Although this is the largest study of automated methods for measuring the size change of pulmonary nodules to date, there are many opportunities for extending this work in the future. The number of nodules and number of methods, though large, was not sufficient to draw statistically significant conclusions on the individual methods or on nodule characteristics; expanding the dataset and gathering results from more automated methods would allow for more statistically significant conclusions. In expanding the dataset, scans from additional manufacturers should be included to determine if that has an effect; all of the scans in this study were acquired on scanners from the same manufacturer (GE Healthcare). Finally, future studies could provide the entire CT scan for analysis – in this study, the CT scans were clipped in the axial direction to both reduce the amount of data to be distributed and because some cases, since this was a retrospective study, did not have CT scans of the entire lung region. Some au-

tomated methods were developed with the assumption that many more image slices would be available, and these methods may perform non-optimally with this data.

2.6 Summary

An important consideration when developing a new algorithm how it performs relative to previous methods. Addressing this has been problematic in the past, due to the use of different datasets and evaluation methods. With this study, a uniform dataset and evaluation methodology have been established to evaluate automated nodule size measurement methods. The framework now exists to both evaluate new methods and to develop further studies.

Size change measurements made on 50 nodule image pairs were reported for 18 different methods. The analysis of the results showed (a) that there was no statistical difference between the methods on scans of the same slice thickness, (b) that there was a statistical difference in the methods when the scan slice thickness is changed, and (c) that the behavior of the methods for nodules with a small real change in size was similar to that for the zero-change data. The last point has implications for the validity of using zero-size change datasets for evaluating nodule measurement performance. Both the techniques used by the methods and the amount of automation did not have a large effect on the performance of individual methods. Finally, the nodules included in the dataset spanned a wide variety of scanner parameters and morphology. In group A, nodules that appeared on differing numbers of slices tended to have larger biases, and in group C, nodules with attachments had higher variation than nodules without attachments. The other nodule characteristics did not have a clear correlation with the amount of variation or bias for the methods.

For 12 of the methods, volume measurements were provided in addition to the size change measurements. The volume measurements did show a statistical difference between methods; therefore, caution is needed when extrapolating from studies that focus only on volume estimation when size change is the intended task.

CHAPTER 3

IMPROVED NODULE VOLUME MEASUREMENT REPEATABILITY USING MOMENT-BASED Z-COMPENSATION

A novel method has been developed to improve volume measurement repeatability by compensating for additional variation in the axial (z) dimension compared to the in-plane (x and y) dimensions [62]. CT images are usually anisotropic, having a lower resolution in the z-dimension compared to the x- and y-dimensions. Furthermore, it is possible that the CT image formation process (acquisition and reconstruction) may result in some increased geometric distortion with respect to the z-dimension, due to additional challenges from the use of helical, multi-detector acquisition that require more complex reconstruction algorithms. This may account for some of the observed change in size of the zero-change nodules in the VOLCANO challenge in spite of the lack of any physical change in the nodule, as discussed in the previous chapter. A z-compensation (ZCOMP) algorithm which is specifically designed to accommodate such variation when estimating nodule size and change in size is presented and evaluated in this chapter. The ZCOMP algorithm is based on a robust estimate of the nodule size using image moments that are invariant to z-dimension linear distortion (lengthening or shortening) but not x or y variations. This method provided a reduction in variability, compared to a volumetric segmentation method, on a dataset of zero-change nodules.

3.1 Z-variation compensation model

The underlying assumption of the z-compensation algorithm is that the measurement in the z-dimension is less reliable than the measurements in the other dimensions. By taking this into consideration when estimating the volume of

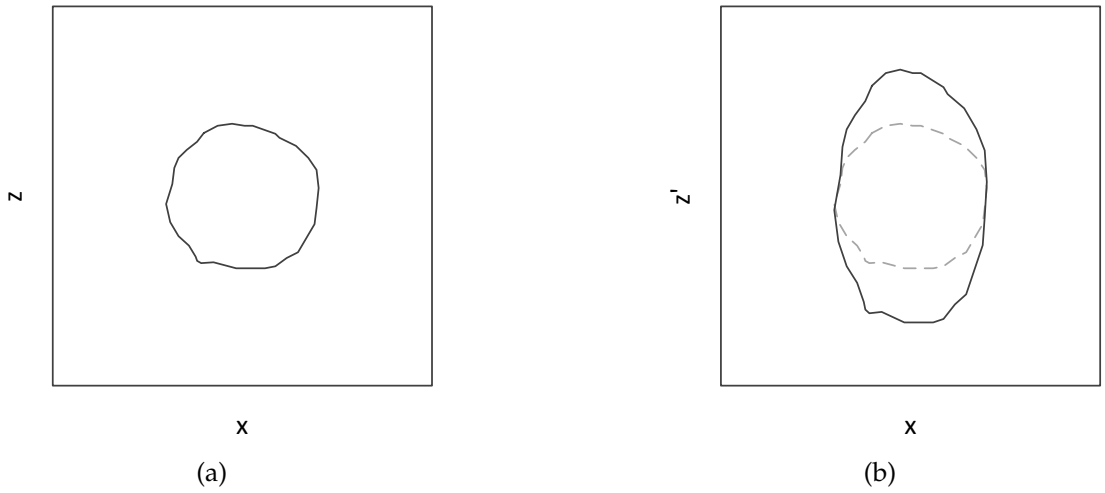


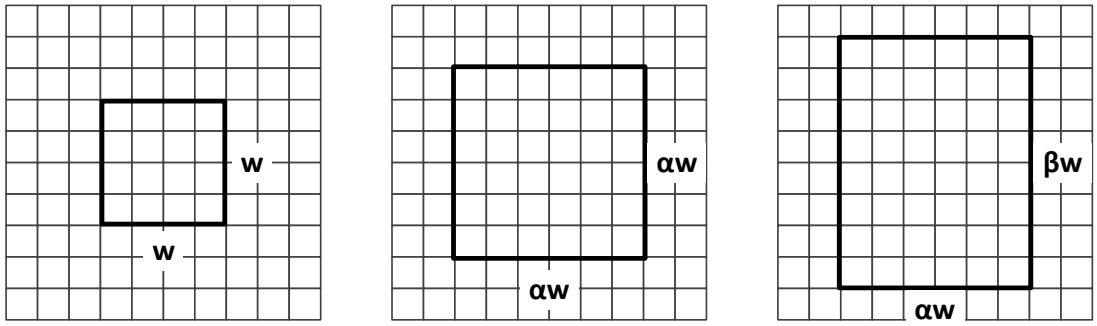
Figure 3.1: Diagram of 2D nodule with a) no distortion and b) small distortion in the z-dimension. The distortion increases the apparent size of the nodule.

a nodule, we can reduce the measurement variability. Consider a 2D image (x and z dimensions) of a nodule region shown in Figure 3.1 for which there exists a small but linear distortion in the z-dimension. To accurately measure the growth in a time-separated pair of such images, we derive a method that compensates for the distortion. Our approach is based on two assumptions:

1. Uniform isotropic growth of the nodule
2. Linear distortion in one dimension

The uniform growth assumption allows for the measurement of the growth of the entire nodule from the change in one of the dimensions, and the assumption of linear distortion provides a measurable model that we can analyze.

Using these two assumptions, a model can be derived. Let us begin with a 2D square that, in one case, grows uniformly in both dimensions by a factor α, but in the second case, grows in one dimension by α and in the other dimension by β. This is illustrated in the digram in Figure 3.2.



(a) (b) (c)

Figure 3.2: Illustration of an object with a) an original size of w^2 that grows b) uniformly to $\alpha^2 w^2$ or c) non-uniformly to $\alpha\beta w^2$

The original area of the square is w^2. It grows uniformly to $\alpha^2 w^2$ and non-uniformly to $\alpha\beta w^2$. In the uniform growth case, the growth Δ_u is:

$$\Delta_u = \frac{A_2 - A_1}{A_1} = \frac{\alpha^2 w^2}{w^2} - 1 = (\alpha^2 - 1)$$

In the non-uniform growth case, the growth Δ_n would be:

$$\Delta_n = \frac{A_2 - A_1}{A_1} = \frac{\alpha\beta w^2}{w^2} - 1 = (\alpha\beta - 1)$$

which differs from the uniform case. We would like the growth to be independent of any distortion in a single dimension, and for this, we use image moments.

3.1.1 Moment analysis of an image region

One method to measure the size of an object in an image is via image moments. Moment analysis of images has been used for various tasks including such diverse applications as shape characterization for the identification of airplanes [63], nodule characterization algorithms [64, 65], and morphological characterization of intracranial aneurysms [66]. The general equation for moments of

order $p + q$ for a two-dimensional image is:

$$m_{pq} = \int_x \int_z x^p z^q f(x, z) \tag{3.1}$$

where x and z are coordinates in space and $f(x, z)$ is the image function. A 2D image is composed of discrete voxels, so the discrete version of Equation 3.1 is:

$$m_{pq} = \sum_x \sum_z x^p z^q f(x, z) \tag{3.2}$$

where x and z are the voxel coordinates and $f(x, z)$ is the intensity of the voxel. For a binary image, $f(x, z)$ is either 0 or 1, which results in the computation of geometric moments – the moments are computed on the geometric form of an object, where only the pixel locations belonging to the object are considered and not the intensity. Note that this function is sensitive to the location of the region under consideration (e.g. if the same region is shifted ten pixels in the positive x-direction, it would have a larger m_{10} moment). This is undesirable since the same object in different locations should, for the purpose of size measurement, result in the same value, assuming no change in size. Location invariance of the moment values is achieved by computing the moments with the index origin at the center of mass (COM) of the binary object region according to the following expression:

$$\mu_{pq} = \int_x \int_z (x - \bar{x})^p (z - \bar{z})^q f(x, z) dx\, dz \tag{3.3}$$

which in the discrete image domain is

$$\mu_{pq} = \sum_x \sum_z (x - \bar{x})^p (z - \bar{z})^q f(x, z) \tag{3.4}$$

where the COM is given by $(\bar{x}, \bar{z})$: $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{z} = \frac{m_{01}}{m_{00}}$. Note that, from the moments, common properties of the image may be derived. For example, the number of voxels in the segmentation is given by $\mu_{00} = \sum_x \sum_z f(x, z)$. The following section will show how image moments can be used for estimating the size of objects.

3.1.2 Compensation for one-dimensional distortion with image moments

Image moments provide a framework that can be used to estimate the size of an object in the presence of a distortion in a single dimension. We shall begin our discussion by analyzing the situation presented earlier in Figure 3.2. The original size of the square, Figure 3.2a, has a central moment μ_{pq_A}. The increase in size of the object in Figure 3.2b can be represented as a scaling of the original image, that is, $f_B(x,z) = f_A(\frac{x}{\alpha}, \frac{z}{\alpha})$. This leads to the expression for the central moment in the presence of scaling:

$$\mu_{pq_B} = \int_x \int_z x^p z^q \, f(\frac{x}{\alpha}, \frac{z}{\alpha}) \, dx \, dz \tag{3.5}$$

which, after substituting $x' = \frac{x}{\alpha}$ and $z' = \frac{z}{\alpha}$ becomes

$$\mu_{pq_B} = \int_x \int_z \left(\alpha x'\right)^p \left(\alpha z'\right)^q \alpha^2 f(x', z') \, dx \, dz = \alpha^{p+q+2} \mu_{pq_A} \tag{3.6}$$

Following the same substitutions, the object with non-uniform growth in Figure 3.2c has the central moment

$$\mu_{pq_C} = \alpha^{p+1} \beta^{q+1} \mu_{pq_A} \tag{3.7}$$

The goal is to have an estimate of the size that will be invariant to the non-uniform growth β. This is accomplished if we define the size to be

$$S_{2D} = \phi \frac{\mu_{20}}{\mu_{00}} \tag{3.8}$$

where ϕ is included as an unknown scaling factor which may not be necessary. Equation 3.8 provides the same size estimates for the the uniform growth case

$$S_{uniform} = \phi \frac{\alpha^4 \mu_{20_A}}{\alpha^2 \mu_{00_A}} = \phi \alpha^2 \frac{\mu_{20_A}}{\mu_{00_A}} \tag{3.9}$$

and the non-uniform case

$$S_{non-uniform} = \phi \frac{\alpha^3 \beta \mu_{20_A}}{\alpha \beta \mu_{00_A}} = \phi \alpha^2 \frac{\mu_{20_A}}{\mu_{00_A}} \tag{3.10}$$

Since the size estimates are the same, the growth estimates Δ for both of these cases are the same as well and are:

$$\Delta = \frac{S_{B_{2D}}}{S_{A_{2D}}} - 1 = \frac{\phi \alpha^2 \frac{\mu_{20_A}}{\mu_{00_A}}}{\phi \frac{\mu_{20_A}}{\mu_{00_A}}} - 1 = \alpha^2 - 1 \tag{3.11}$$

which also matches the growth computed from the area A of the rectangles, assuming uniform growth:

$$\Delta_u = \frac{A_B}{A_A} - 1 = \frac{\alpha w \cdot \alpha w}{w \cdot w} - 1 = \alpha^2 - 1 \tag{3.12}$$

In calculating growth, the scaling factor ϕ can be eliminated because it is present in both the numerator and denominator, but this does not mean that the size S is equivalent to the size that would be computed from the dimensions of the object. The scaling factor ϕ can be determined from the moment expressions (the object is assumed to be centered at the origin so that the COM is located at (0,0), and the object is bounded from $-\frac{w}{2}$ to $\frac{w}{2}$):

$$\mu_{20} = \int_{-w/2}^{w/2} \int_{-w/2}^{w/2} x^2 \, dx \, dy = \int_{-w/2}^{w/2} \frac{1}{3} \left(\left(\frac{w}{2} \right)^3 - \left(-\frac{w}{2} \right)^3 \right) \, dy = \frac{w^3}{12} w = \frac{w^4}{12}$$

$$\mu_{00} = \int_{-w/2}^{w/2} \int_{-w/2}^{w/2} dx \, dy = \int_{-w/2}^{w/2} w \, dy = w^2$$

so that, for S to be equivalent,

$$\phi = 12$$

This two-dimensional moment-based measure of size makes use of all the available image data and is unaffected by a distortion in a single dimension.

The above analysis can be extended to three-dimensions. In 3D, we assume that the measurement in the z-dimension may have distortion, while the mea-

surements in the x- and y-dimensions are reliable. Again, we use moment analysis to compute the size of the object, and the equation for a discrete, central 3D moment is:

$$\mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r f(x, y, z) \tag{3.13}$$

where x, y, and z are the voxel coordinates and $f(x, y, z)$ is the intensity of the voxel, and the COM is given by $(\overline{x}, \overline{y}, \overline{z})$: $\bar{x} = \frac{m_{100}}{m_{000}}$, $\bar{y} = \frac{m_{010}}{m_{000}}$, and $\bar{z} = \frac{m_{001}}{m_{000}}$. As in the two-dimensional case, the size in a single dimension is based on the second-order central moment of the reliable dimensions, the x- and y- dimensions. The size is computed now using the size in x and y, so the size is given by:

$$S_{3D} = k \left(\sqrt{\frac{\mu_{200}}{\mu_{000}}} \cdot \sqrt{\frac{\mu_{020}}{\mu_{000}}} \right)^{3/2} \tag{3.14}$$

where k is a scaling factor, $\sqrt{\frac{\mu_{200}}{\mu_{000}}}$ is an estimate of the size in a single dimension, since S_{2D} provides a two-dimensional size estimate, and the cube of the square root of the product is done to compute the volume from measures in two dimensions.

Having derived the expression for the 3D size (volume) that is independent of the distortion in the z-dimension, we can now develop an algorithm to measure the size and size change of pulmonary nodules.

3.2 Pulmonary nodule growth measurement with Z-compensation

Based on the z-variation compensation model described in Section 3.1, an algorithm was developed to measure pulmonary nodule growth. In a typical nodule growth measurement algorithm, such as the algorithm described in Section 1.7, the nodule is segmented on each scan, the volume computed by summing the number of voxels in the segmented image of the nodule and multiplying by the

voxel size to obtain the volume of the nodule, and estimating the growth rate from the volume change. The z-variation compensation algorithm (ZCOMP), on the other hand, calculates a surrogate measure of size based on the three-dimensional moments of the image. ZCOMP has three primary steps:

1. Given a seed point within the nodule, segment the nodule on each scan to obtain a binary representations of the nodule
2. Compute the zero and second order moments, μ_{200} and μ_{020}, on each binary image of the nodule
3. Estimate growth based on the moment analysis of the nodule

Obtaining a binary representation of the nodule from a user-specified seed point is accomplished by segmenting the nodule using the algorithm described in Section 1.7 by Reeves et al. [23]. The algorithm provides a binary image representation of the nodule, from which the 3D moments are computed, as described in Section 3.1.1. Finally, these moments are used to compute a surrogate growth measure, which is further described in the next section.

3.2.1 Growth analysis from moments

A surrogate measure of pulmonary nodule growth is computed from the 3D moment analysis of the binary image representation of the nodule. The quantity of interest for growth analysis is the relative size difference:

$$\text{RSD} = \frac{S_{T2}}{S_{T1}} - 1$$

where S_2 and S_1 are the measured sizes (such as volume) on the second ($T2$) and first ($T1$) scans respectively. Substituting Equation 3.14 into the above ex-

pression yields the z-compensated relative size difference:

$$\mathrm{RSD} = \frac{\left(\sqrt{\frac{\mu_{200}}{\mu_{000}}} \cdot \sqrt{\frac{\mu_{020}}{\mu_{000}}}\right)^{3/2}_{\mathrm{T2}}}{\left(\sqrt{\frac{\mu_{200}}{\mu_{000}}} \cdot \sqrt{\frac{\mu_{020}}{\mu_{000}}}\right)^{3/2}_{\mathrm{T1}}} - 1$$

3.2.2 Derivation of scaling factor k

To compute the actual volume of the nodule, Equation 3.14 needs to be scaled by a constant k, but since we are concerned with the relative size change when calculating growth, k cancels out of the expression and we need not be concerned with its value. However, to compare the volume estimates of ZCOMP to the volume estimates of the segmentation method, we can derive k. For this derivation, we will assume that, since nodules are roughly spherical, that the shape will be a sphere of radius r. The continuous expression for the moment will be used for the derivation. The equation for a sphere is:

$$x^2 + y^2 + z^2 = r^2 \tag{3.15}$$

for a sphere of radius r. The continuous moment μ_{200} is:

$$\mu_{200} = \int_x \int_y \int_z x^2 f(x, y, z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{3.16}$$

To get the limits of the integral, we rearrange equation 3.15

$$z = \sqrt{r^2 - x^2 - y^2} \tag{3.17}$$

and, to simplify the problem, we will only compute the integral for a quarter of a sphere, so the limits on z are $0 \leq z \leq \sqrt{r^2 - x^2 - y^2}$. To determine the limits on y, the projection of the sphere onto the xy plane is

$$y = \sqrt{r^2 - x^2}$$

and again, since the integral is being computed for a quarter of a sphere, the limits on y are $0 \leq y \leq \sqrt{r^2 - x^2}$. Finally, the limits on x are $-r \leq x \leq r$. Substituting these limits into equation 3.16 results in the integral

$$\mu_{200} = 4 \int_{-r}^{r} \int_{0}^{\sqrt{r^2-x^2}} \int_{0}^{\sqrt{r^2-x^2-y^2}} x^2 \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

since $f(x, y, z)$ is 1 within the integral. Solving the integral results in the following:

$$\mu_{200} = 4 \cdot \frac{\pi}{15} r^5$$

Now we can substitute into equation 3.14, using the volume of a sphere as V, and solve for k:

$$\begin{aligned}
\frac{4}{3}\pi r^3 &= k \left(\sqrt{\frac{\frac{4\pi r^5}{15}}{\frac{4}{3}\pi r^3}} \cdot \sqrt{\frac{\frac{4\pi r^5}{15}}{\frac{4}{3}\pi r^3}} \right)^{3/2} \\
&= k \left(\sqrt{\frac{r^2}{5}} \cdot \sqrt{\frac{r^2}{5}} \right)^{3/2} \\
&= k \left(\frac{r^2}{5} \right)^{3/2} \\
&= k \left(\frac{r^3}{5\sqrt{5}} \right)
\end{aligned}$$

Rearranging the equation to solve for k:

$$\begin{aligned}
\frac{4}{3}\pi r^3 &= k \left(\frac{r^3}{5\sqrt{5}} \right) \\
\frac{20\sqrt{5}}{3}\pi &= k
\end{aligned} \tag{3.18}$$

This scaling coefficient was validated using synthetically generated spheres. A VisionX program, `vgsphere`, was used to generate spheres ranging in diameter from 4.0 mm to 8.0 mm. The `vgsphere` program is capable of generating spheres with partial voxel intensities. The output of vgsphere was thresholded

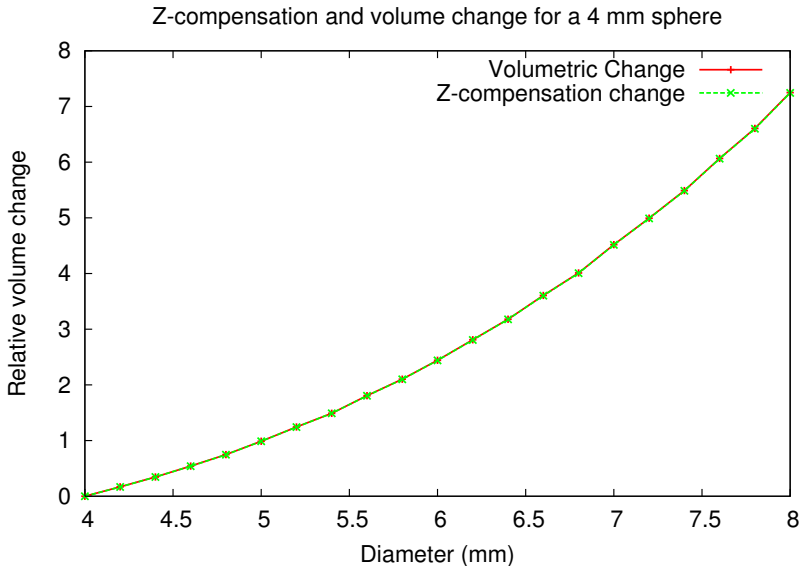


Figure 3.3: Plot of volume change of a 4.0 mm synthetic sphere using the volumetric and z-compensation algorithms; the x-axis indicates the new diameter of the sphere used to calculate S_{T2}. In this synthetic case, both algorithms measure the same amount of change.

to the same value as in the segmentation method described in Section 1.6, and the volume change was computed relative to a 4.0 mm diameter sphere. The estimated volume change for the volumetric and z-compensation algorithms are shown in the plot in Figure 3.3, with the residuals shown in Figure 3.4. For an increase in nodule diameter to 5.0 mm, which corresponds to a relative volume change of 0.988 or 98.8%, the z-compensation algorithm measured a change of 0.987 or 98.7%, a difference of only 0.1%. This small difference is attributable to the imprecision from rounding errors in the sphere generation and floating point representation and is not significant. ZCOMP was within 1% of the volumetric method up to a relative volume change of 7.25 (725%), and there was a trend of increasing difference with increasing relative volume change, which is negligible.

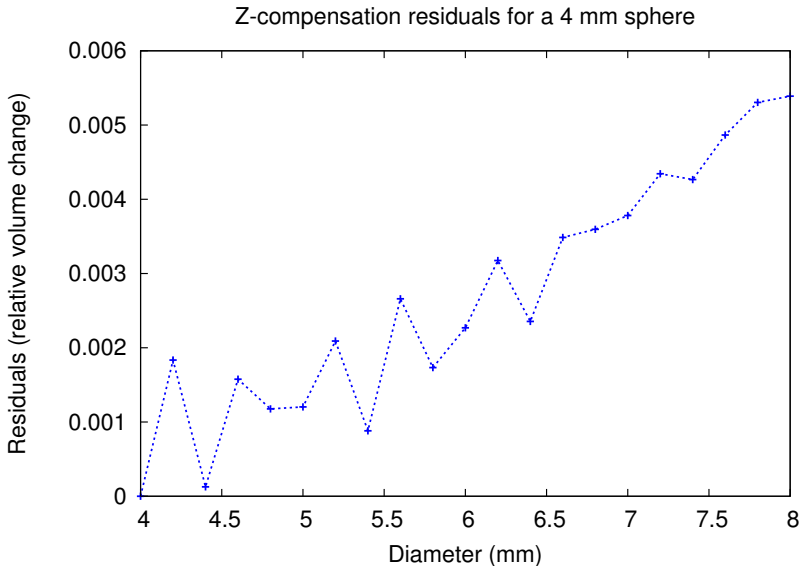


Figure 3.4: Plot of residual relative volume change of the z-compensation algorithm compared to the volumetric method on the baseline 4.0 mm diameter sphere from Figure 3.3. These small residuals do not have any noticeable impact on the measurement.

3.3 Experimental evaluation of z-compensation

The goal of the z-compensation method was to reduce the variability from the additional uncertainty in the z-dimension while still being sensitive to actual changes in the nodule size. An experiment to quantify this was developed using nodules with repeat scans in the same session—since these nodules should have no change, any non-zero measured change would be due to problems with the scans or uncertainties in the measurement method. The relative size differences (RSD) were measured for these nodules, and lower values of RSD indicate less variation. The computed RSD were used to compute the interscan variability, which is defined as the 95% confidence interval of the RSD. Given the mean, μ, and standard deviation, σ, the 95% confidence interval of the RSD is:

$$(\mu - 1.96 \cdot \sigma,\ \mu + 1.96 \cdot \sigma)$$

To evaluate the method, the RSD was computed for a dataset of 22 nodules

with multiple scans taken within the span of a few minutes. To determine the effect of varying slice thicknesses on the performance of the method, these 22 nodules were divided into two subsets of eleven nodules each. In the first set of eleven cases, the scans were taken using the same slice thickness, with ten of the eleven cases at 1.25 mm slice thickness and one at a slice thickness of 2.5 mm. In the remaining eleven cases, scans were taken at different resolutions, with at least one scan at a slice thickness of 1.25 mm. In three cases, the second scan had a 5.0 mm slice thickness while eight had 2.5 mm. All cases had a nodule of solid consistency, as determined by a radiologist, with at least two scans that included the entire nodule. The mean size of the nodules in the dataset was 14.1 mm with a standard deviation (SD) of 5.6 mm. For the eleven cases with scans of the same slice thickness, the mean size was 14.8 mm with a SD of 4.6 mm while the eleven cases with scans of different slice thickness had a mean size of 13.4 mm with a SD of 6.5 mm. A plot of the size distribution of the nodules in the dataset is shown in Figure 3.5. Scans were obtained using either a GE LightSpeed QX/i or LightSpeed Ultra scanner, using 120 kVp and a current in the range of 40-250 mAs.

The standard volumetric method and ZCOMP were compared on the basis of interscan variability, with a smaller interscan variability indicative of a more consistent measurement.

3.4 ZCOMP evaluation results

The new ZCOMP method was compared to a previously published volumetric method on the basis of interscan variability by Reeves et al. [23] described in Section 1.6. Interscan variability is presented as the upper and lower bounds

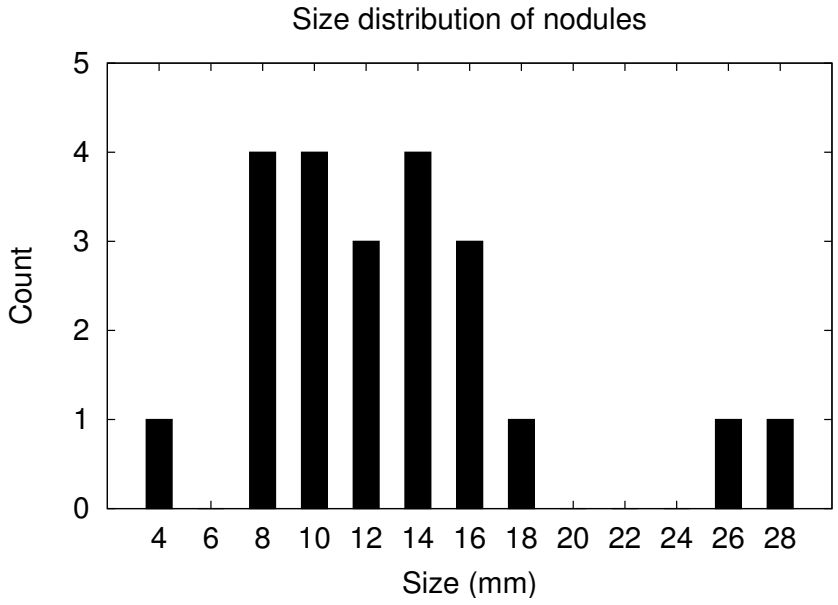


Figure 3.5: Size distribution of nodules in the ZCOMP evaluation dataset. The value along the x-axis indicates the minimum size of nodules in the bin.

Table 3.1: Interscan variability of volumetric and moment-based methods

Dataset	Interscan variability (%)	
	Volumetric method	Moment-based method
Full (22 nodules)	(-52.1, 30.1)	(-34.2, 23.3)
Same slice thickness (11)	(-24.0, 18.2)	(-12.4, 12.7)
Mixed slice thickness (11)	(-68.4, 30.2)	(-46.5, 24.4)

of the 95% confidence interval of the relative size difference of the cases in the study. On cases with scans of the same slice thickness, which is comparable to cases used in previously published studies, the volumetric method had an interscan variability of -24.0% to 18.2%. The moment-based method had a lower interscan variability of -12.4% to 12.7%. The full results are presented in Table 3.1, and were previously presented at SPIE Medical Imaging 2009 [62].

As the conventional method of expressing pulmonary nodule size is in diameter, the interscan variability can be converted to a relative size difference; these values are included in Table 3.2. As an example of the actual impact on size measurement the interscan variability may have, the size difference corresponding to the confidence intervals of the interscan variability in Table 3.1 are shown in Table 3.3 for a 10 mm nodule. Note that despite a large interscan variability in relative volume difference, the difference in size on a 10 mm lesion is less than 1 mm for cases of the same slice thickness.

3.5 ZCOMP Discussion

In the ideal case, a nodule growth measurement method would have 0% variability on a zero-change scan, since the nodule is not changing in size. Lower

Table 3.2: Interscan variability presented as relative size difference

Dataset	Interscan variability (size) (%)	
	Volumetric method	Moment-based method
Full (22 nodules)	(-21.8, 9.2)	(-13.0, 7.2)
Same slice thickness (11)	(-5.7, 8.7)	(-4.1, 4.3)
Mixed slice thickness (11)	(-31.9, 9.2)	(-18.8, 7.5)

Table 3.3: Interscan variability presented as size range for a 10 mm nodule

Dataset	Interscan size variability (mm)	
	Volumetric method	Moment-based method
Full (22 nodules)	(7.8, 10.9)	(8.7, 10.7)
Same slice thickness (11)	(9.1, 10.6)	(9.6, 10.4)
Mixed slice thickness (11)	(6.8, 10.9)	(8.1, 10.8)

values of interscan variability indicate a more reliable volume measurement, which is critical when computing the volume change. This variability can then be used to establish error bounds on the measurement precision of the method. Previous studies of other methods have found interscan variability of approximately (-20%, 20%) using scans of the same slice thickness, as summarized in Section 1.4.4; our standard volumetric method, on the eleven nodules with scans of the same slice thickness, performed similarly with a variability of (-24.0%, 18.2%). ZCOMP reduced the variability to (-12.4%, 12.7%). As described in Section 1.4, by reducing the measurement variation, we can compute more accurate growth estimates, in turn improving clinical decision making.

Often, scans of a patient may have been taken at different resolutions due to changes in protocol or scanners. To assess performance for such a case, this study included zero-change scans with different slice thicknesses. Both methods performed worse on the eleven nodules with scans of different slice thicknesses, with the volumetric method interscan variability increasing to (-68.4%, 30.2%) and the moment-based method interscan variability increasing to (-46.5%, 24.4%). Although ZCOMP had better performance than the volumetric method, the intervals for both methods are still too large to be useful. This result suggests that pulmonary nodule growth assessment requires scans with the same slice thickness for the most accurate measurement.

Although the measurements from ZCOMP had a lower interscan variability overall, the volume difference measured by the algorithm was not consistently lower on every case. These observations indicate that ZCOMP is not simply less sensitive to changes in general. The z-compensation algorithm measured a higher relative volume difference than the volumetric method on ten cases, but the difference between the two methods was small, with an average difference between the two methods of 5.8%. These cases were divided nearly equally divided among the same-slice thickness (4) and mixed-slice thickness (6) cases and were generally the cases with small volume differences. On the other twelve cases where ZCOMP measured a lower relative volume difference than the volumetric method, ZCOMP had much lower values; on average, ZCOMP had measurements that were 14.0% lower. These cases tended to be those nodules with large volume differences, where ZCOMP was closer to the ideal measurement of 0.

All of these nodules were zero-change nodules, yet most of the nodules had measured non-zero change. For many of the nodules, the appearance of the nodule on one scan was different than on the subsequent scan; in some cases, the difference was great enough to cause a large difference in volume between the two scans, while in others the volume difference was minimal. One case with a large difference in volume is shown in Figure 3.6. In this case, one scan was acquired with a slice thickness of 1.25 mm while the other scan was acquired with 5.0 mm; the relative volume difference computed by the volumetric method was 37.8% while the relative size difference computed by ZCOMP was 43.8%, with the larger volume measured on the first scan. In another case, even though the scans were obtained at different slice thicknesses, the volume change

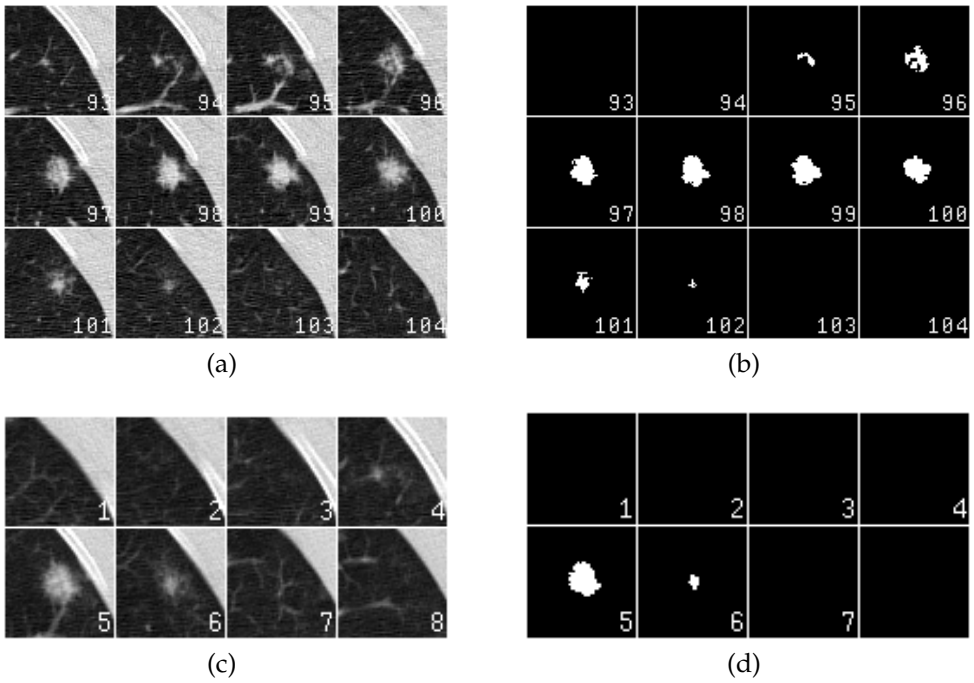


Figure 3.6: Example of a case with scans of the different slice thickness (1.25 mm top, 5.0 mm bottom) and high interscan variability. a) Montage of several slices through the nodule on the first scan, b) segmentation of the nodule on the first scan where white voxels are those belonging to the nodule, c) several slices on the second scan, and d) segmentation on the second scan. Scans are not to the same scale.

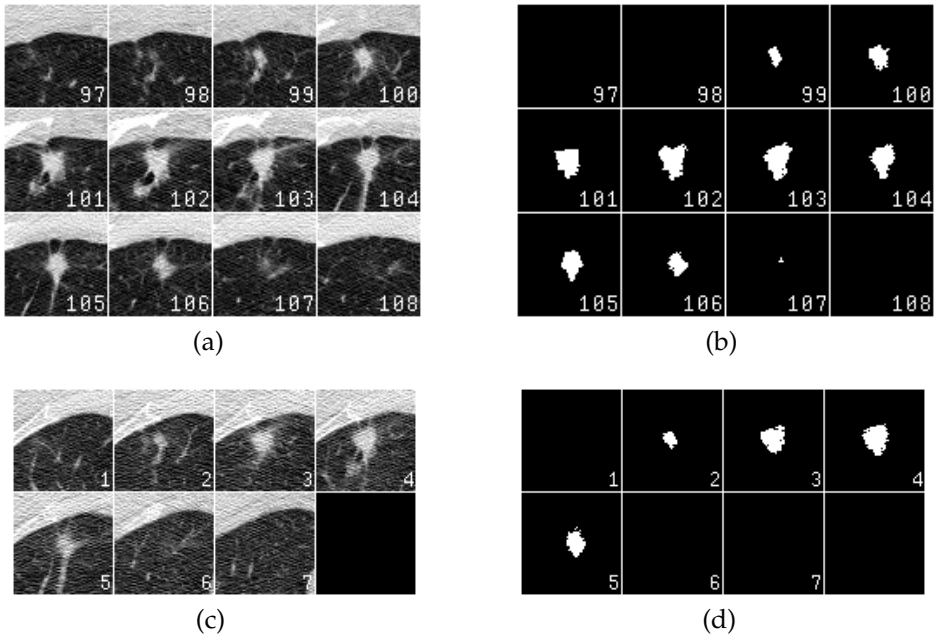


Figure 3.7: Example of a case with scans of the different slice thickness (1.25 mm top, 2.5 mm bottom) and low interscan variability. a) Montage of several slices through the nodule on the first scan, b) segmentation on the first scan with voxels part of the nodule indicated in white, c) several slices on the second scan, and d) segmentation on the second scan. Scans are not to the same scale.

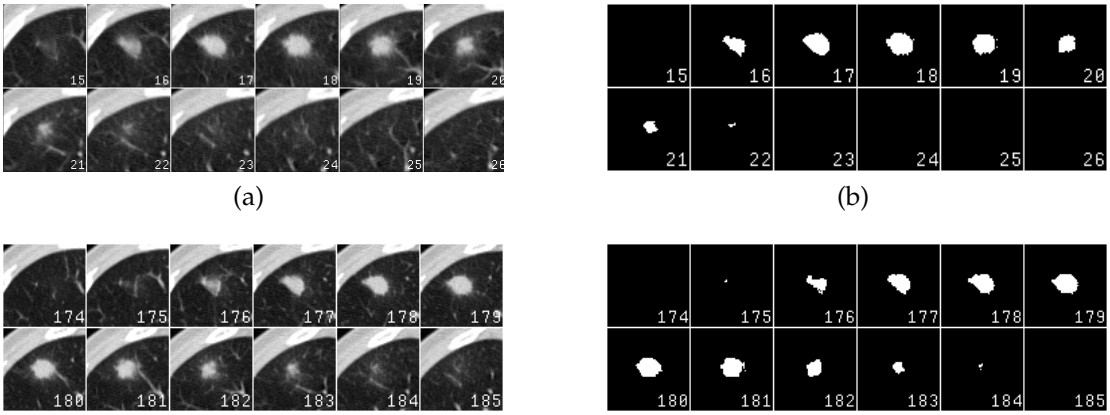
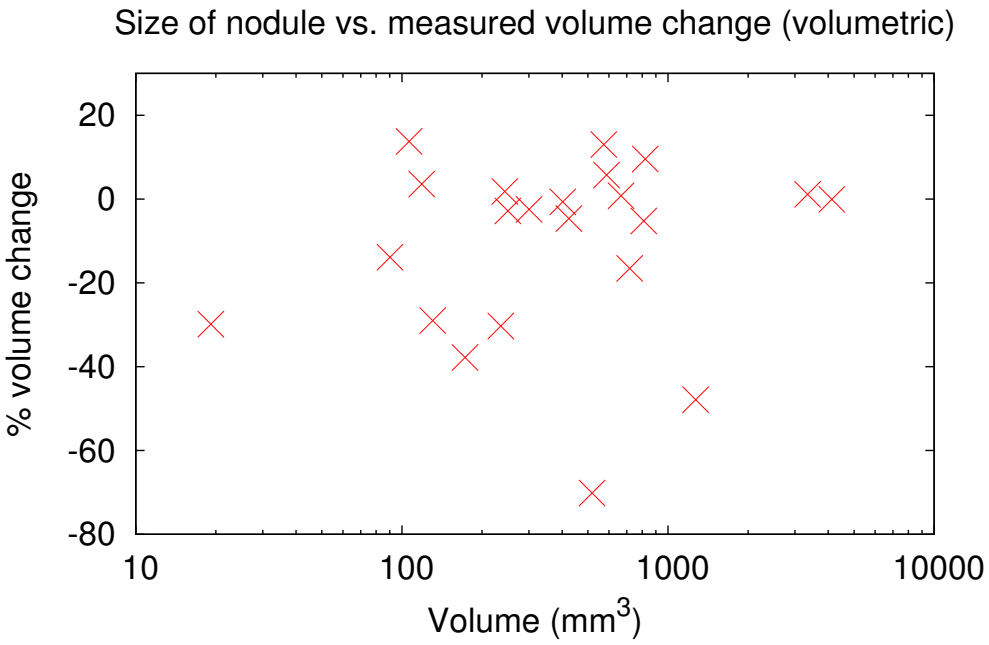


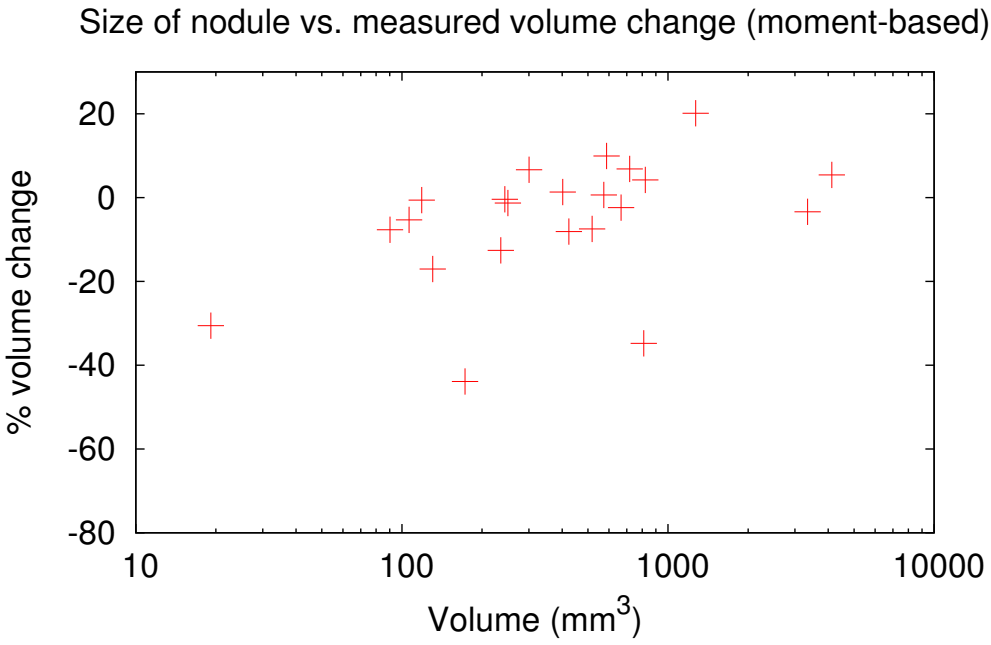
Figure 3.8: Example of a case with scans of the same slice thickness and high inter-scan variability. a) Montage of several slices through the nodule on the first scan, b) segmentation on the first scan, c) several slices on the second scan, and d) segmentation on the second scan. Note that there appears to be an extra slice between the two scans.

was small, as in Figure 3.7. For this case, the volumetric method had a volume difference of 2.8% compared to 13% for ZCOMP. Even in cases with scans of the same slice thickness, there was sometimes a marked difference in the appearance of the nodule between the two scans, as illustrated by the case in Figure 3.8. The nodule appears on an additional slice in the second CT scan, which is reflected in the segmentation shown on the right side of the figure. In this case, ZCOMP was less affected by the additional slice than the volumetric method, with a volume difference of 12.6% for ZCOMP and 30.3% for the volumetric method.

Another possible source of some of the observed variations may come from the sizes of the nodules themselves, which may have resulted in a relationship between the size of the nodule and the measured size difference. Linear regression was performed to determine if there was any correlation between the volume and measured volume change; neither method showed a significant re-



(a)



(b)

Figure 3.9: Scatter plot of measured volume change versus the initial volume of the nodule for a) volumetric measurement method and b) moment-based method. Neither method has a significant dependency on the size of the nodule.

lationship, with the volumetric method having a Spearman's rank correlation coefficient $\rho = 0.14$ and ZCOMP having a $\rho = 0.51$. This can be seen in a scatter plot of the data in Figure 3.9. Based on the lack of correlation between size and volume change, neither method has a variation dependence on the size of the nodule, which agreed with previously published studies [46].

Overall, the results of the z-compensation algorithm showed a promising reduction in interscan variability compared to the volumetric method. This decrease has the potential to allow for better decision making by radiologists, and the algorithm can be applied to any segmentation-based measurement method.

CHAPTER 4

NODULE GROWTH RATE MEASUREMENT FROM DENSITY CHANGE

The majority of current systems for measuring pulmonary nodule growth rate compute the growth rate by segmentation-based volume measurement methods. These methods segment the nodule from each computed tomography (CT) scan and calculate the volume from the segmented image. The process of segmentation entails assigning a class to each voxel: nodule or non-nodule. However, this task is difficult for many nodules with complex shapes where there is uncertainty as to whether a voxel should be considered part of the nodule.

To address this issue, an alternative approach was developed to measure nodule growth that takes into account the radiodensity of the nodule and does not require explicit segmentation [61]. Instead, this method uses the mean density change of a fixed size region containing the nodule as a surrogate measure of nodule growth. The mean density change is computed after applying a weighting function to reduce the influence of non-nodule structures far from the nodule center. This method has the advantage of being able to measure nodules with a complex appearance that may cause a segmentation-based method to fail, such as the nodule in Figure 4.1. This method was evaluated by comparing performance with our direct volumetric segmentation method. To accomplish this evaluation, we used three nodule image databases consisting of: a) zero-change, b) malignant, and c) benign nodules.

4.1 Pulmonary Nodule Density Change Model

To provide a basis for the measurement of nodule growth rate from the density change, we define a model for the nodule that consists of the following:

- The mean density in a region is proportional to the amount of cells in that

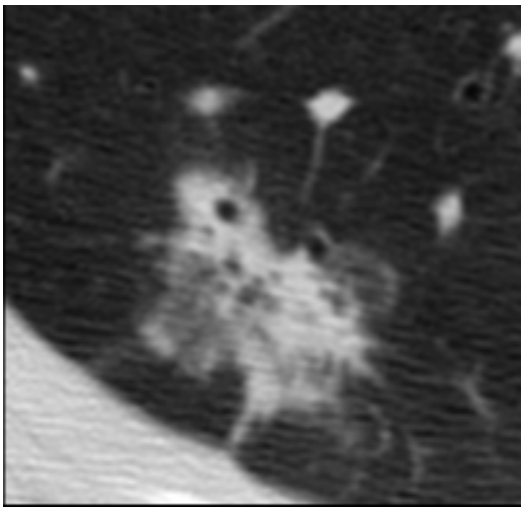


Figure 4.1: Central slice of a nodule exhibiting a complex shape which may be difficult to segment

region

- Nodules have an approximately spherical shape
- The majority of growth occurs at the boundary of the nodule and lung parenchyma
- Non-nodule structures tend to be located at the periphery of the nodule and do not change in size over time

Computed tomography scans have calibrated pixel values that are proportional to the photon density at that location. At the microscopic sub-pixel level, the material in the lung may be classified into two basic types with different densities – air and soft tissue. If we scale the intensity of the image such that the value for air is 0 and the value for soft tissue is 1, then the density of the region of interest (ROI) will indicate the portion of the ROI that is soft tissue. On a subsequent scan, when a nodule has increased in size (by adding more cells and displacing air), the measured density of the same ROI will also increase.

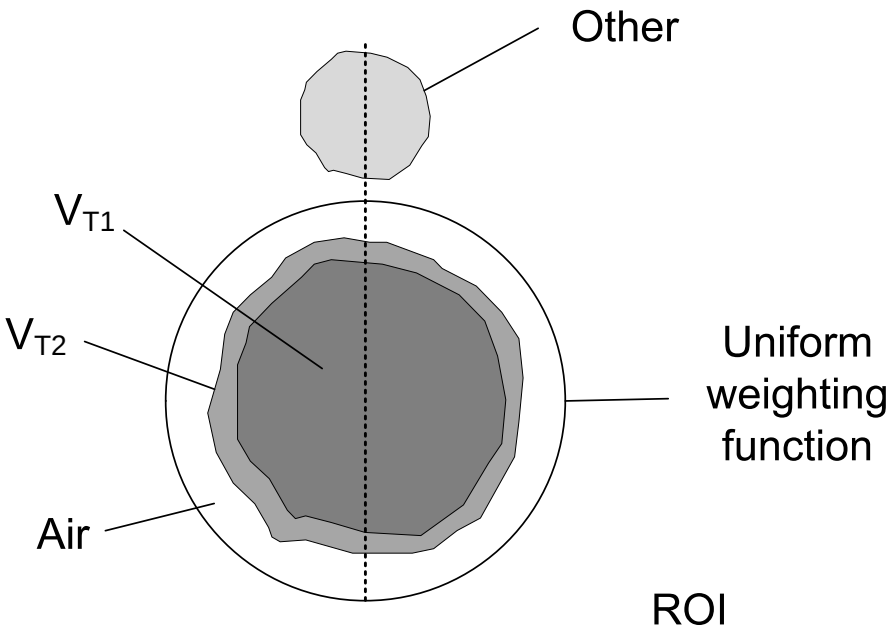


Figure 4.2: Simple model of an isolated nodule surrounded by air. The volume of the nodule grows from V_{T1} to V_{T2}.

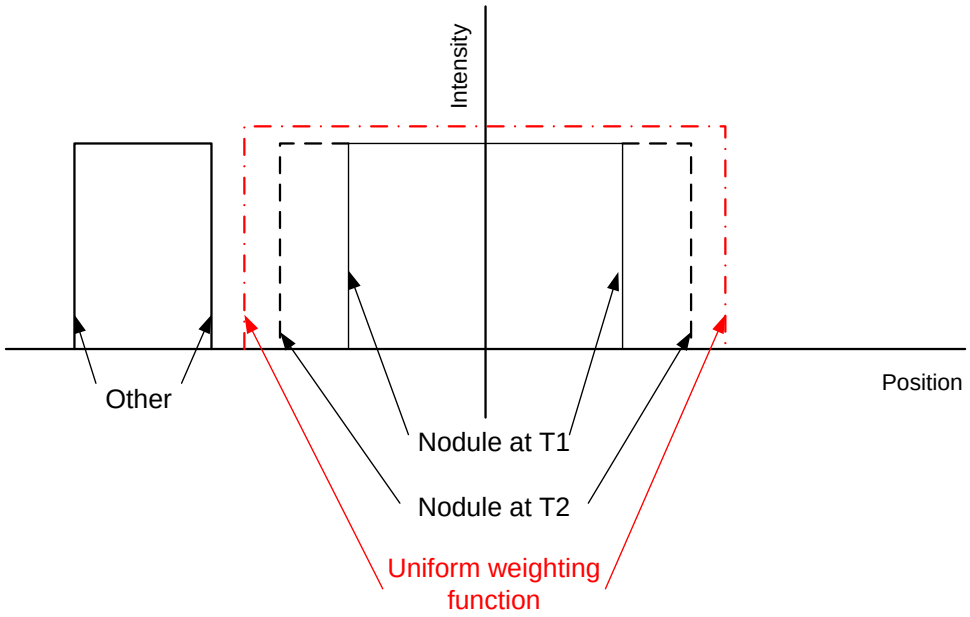


Figure 4.3: Sampled intensities along a line for the simple, isolated nodule model

Let us begin with the simple case of an isolated nodule surrounded only by air and a non-attached, high-intensity structure, as illustrated by the two dimensional region of interest in Figure 4.2. The initial volume of the nodule is V_{T1} on the first scan and increases to V_{T2} on the second scan. The intensity of air is given the value 0, while the voxels of the nodule have an intensity of 1, and the shades of gray only serve to indicate different regions, not intensities. The other structure in the scan is comprised of soft-tisue with an intensity of 1 and is separate from the nodule.

While the mean density could be computed over the entire region of interest, it would be desirable to exclude as many non-nodule voxels as possible to ensure that the mean density change will reflect only the change in the nodule. To accomplish this, we introduce the idea of a *weighting function*, $w(x, y, z)$, which applies a weight to every voxel in the image. In the simplest case, a uniform spherical weighting function could be used that weights voxels located within the sphere as 1 and voxels outside the sphere as 0:

$$w_{\text{uniform}}(x, y, z) = \sum_{x=-r}^{r} \sum_{y=-\sqrt{r^2-(x-x_0)^2}}^{\sqrt{r^2-(x-x_0)^2}} \sum_{z=-\sqrt{r^2-(x-x_0)^2-(y-y_0)^2}}^{\sqrt{r^2-(x-x_0)^2-(y-y_0)^2}} 1$$

where the limits are derived from the equation for a sphere centered at x_0, y_0, z_0, $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$. In Figure 4.2, the weighting function is illustrated as a circle co-located with the nodule center, with a radius large enough to encompass the entire nodule. The radius of the circle is shown to be larger than the nodule to account for uncertainty in the estimation of the nodule size. Sampling the intensities along the dashed line in Figure 4.2 yields the plot shown in Figure 4.3, with the x-axis representing the distance along the line.

The uniform weighting function includes some voxels of air on both sides of the nodule, which represent the uncertainty in the nodule size. Note that including regions of air will not affect the mean weighted density. The mean

weighted density is then given by

$$D = \frac{1}{N_{\text{ROI}}} \sum_x \sum_y \sum_z w(x,y,z) \cdot f(x,y,z)$$

where N_{ROI} is the number of voxels in the ROI and $f(x,y,z)$ is the voxel intensity (1 for soft tissue and 0 for air), and $w(x,y,z)$ is the weighting function, which for this example is the uniform spherical weighting function given above. This simplifies to

$$D = \frac{N_{\text{nodule}}}{N_{\text{ROI}}}$$

where N_{nodule} is the number of voxels in the nodule. Thus, the relative density change (C_D) of the nodule from the first scan to the second scan is

$$C_D = \frac{D_{T2} - D_{T1}}{D_{T1}} = \frac{\frac{N_{T2}}{N_{\text{ROI}}} - \frac{N_{T1}}{N_{\text{ROI}}}}{\frac{N_{T1}}{N_{\text{ROI}}}} = \frac{N_{T2} - N_{T1}}{N_{T1}}$$

where N_{T2} and N_{T1} are the number of voxels comprising the nodule on the second and first scans respectively. If the nodule grew by α between the first and second scans ($N_{T2} = \alpha N_{T1}$), then the C_D is

$$C_D = \frac{\alpha N_{T1} - N_{T1}}{N_{T1}} = (\alpha - 1) \tag{4.1}$$

This can be compared to the relative volume change (C_V) given by

$$C_V = \frac{V_{T2} - V_{T1}}{V_{T1}}$$

where V_{T2} and V_{T1} are the volumes of the nodule on the second and first scans. Note that the volume of the nodule is the volume of a voxel, s, multiplied by the number of voxels. If we again assume the nodule grew by α, then the RVC simplifies to:

$$C_V = \frac{s\alpha N_{T1} - sN_{T1}}{sN_{T1}} = (\alpha - 1) \tag{4.2}$$

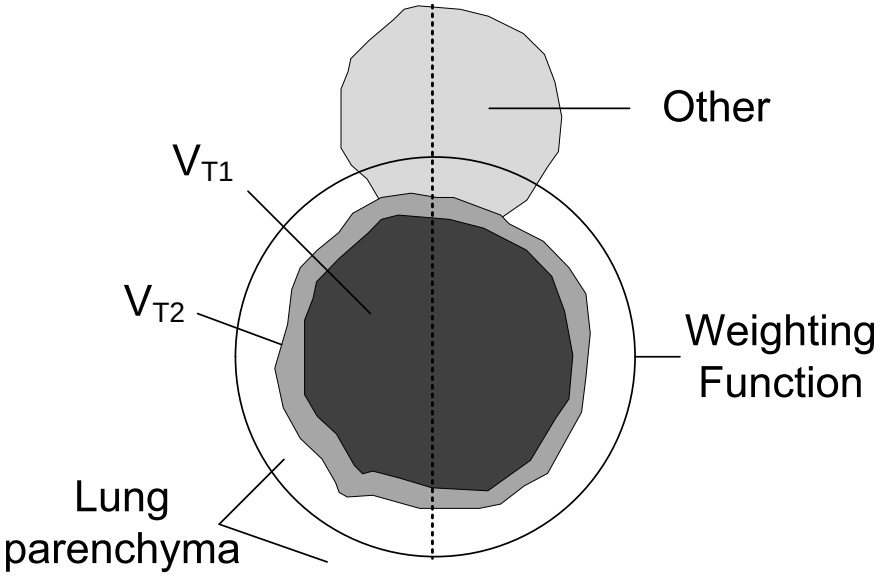


Figure 4.4: Model of nodule region with an attached structure surrounded by air. The nodule increased in size from V_{T1} to V_{T2}.

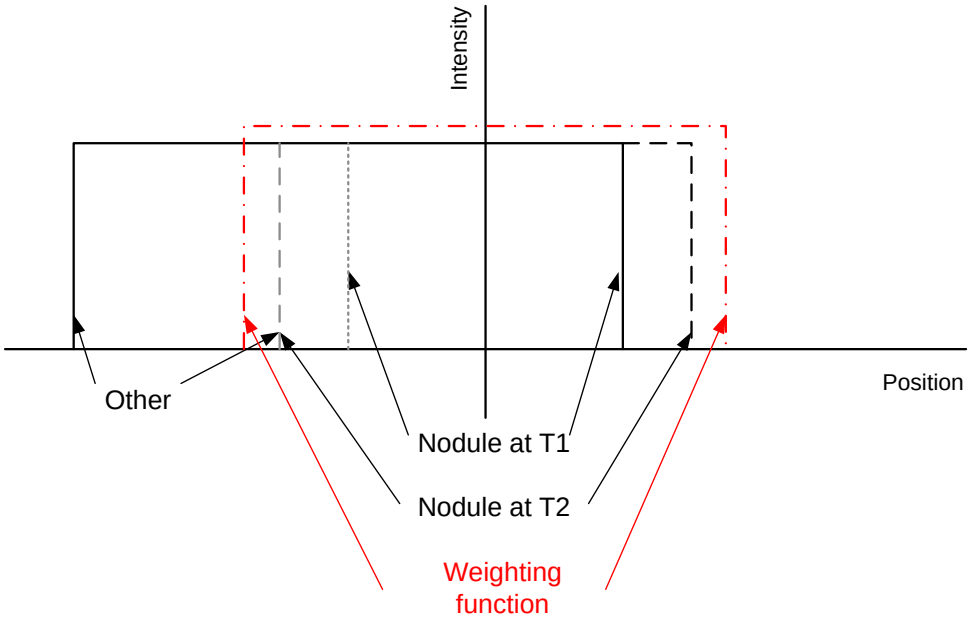


Figure 4.5: Sampled intensities along line for a nodule with an attached structure. The ideal weighting function from Figure 4.3 is shown by the dashed line in red. Note that the boundaries of the nodule and the attached structure can not be determined from the intensity alone; the gray dashed lines only serve for reference.

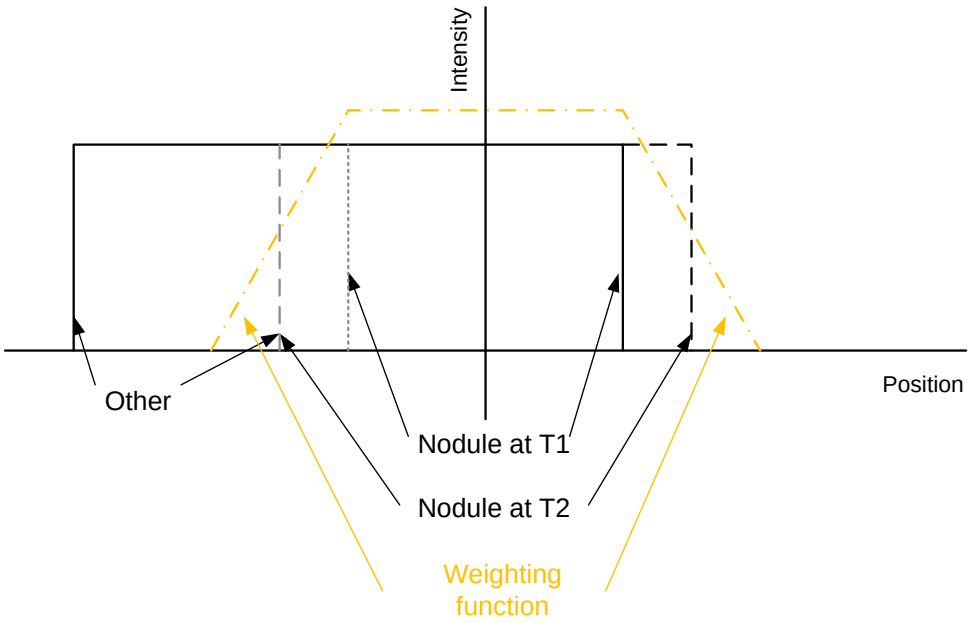


Figure 4.6: Sampled intensities along a line for a nodule with an attached structure. The weighting function, shown by the gray (orange) dotted dashed line, de-emphasizes voxels far from the nodule center.

which is the same expression as the RDC for this case of an isolated nodule.

Many nodules have attached structures, such as blood vessels or the chest wall, that are similar in intensity to the nodule tissue; to model this situation, let us add an attached structure to the ROI and determine its effect on the RDC. In Figure 4.4, the nodule now has a large attached structure. Since the structure is of similar intensity as the nodule, it will be difficult to separate the nodule from the attachment. This is illustrated by the plot in Figure 4.5, which represents the intensities along the dashed line in Figure 4.4. If we use the ideal weighting function from Figure 4.3, some of the voxels from the attached structure will be included in the mean weighted density calculation, due to the uncertainty in the nodule size. Using the same parameters as for the simple model without an attachment, assume that the weighting function includes some number of voxels from the attached structure proportional to the number of voxels in N_{T1},

$N_{\text{att}} = \beta N_{T1}$, and that the number of voxels in the attached structure within the ROI does not change between the first scan and the second scan. The density on the first scan becomes

$$\hat{D}_{\text{T1}} = \frac{N_{\text{T1}} + N_{\text{att}}}{N_{\text{ROI}}} = \frac{N_{\text{T1}} + \beta N_{\text{T1}}}{N_{\text{ROI}}} = D_{\text{T1}} \cdot (1 + \beta)$$

and on the second scan

$$\hat{D}_{\text{T2}} = \frac{N_{\text{T2}} + N_{\text{att}}}{N_{\text{ROI}}} = \frac{\alpha N_{\text{T1}} + \beta N_{\text{T1}}}{N_{\text{ROI}}} = D_{\text{T1}} \cdot (\alpha + \beta)$$

The resulting relative density change, $\hat{C_D}$, with the inclusion of non-nodule voxels, is given by:

$$\hat{C_D} = \frac{\hat{D}_{\text{T2}} - \hat{D}_{\text{T1}}}{\hat{D}_{\text{T1}}} = \frac{D_{\text{T1}} \cdot (\alpha + \beta) - (D_{\text{T1}} \cdot (1 + \beta))}{D_{\text{T1}} \cdot (1 + \beta)} = \frac{\alpha - 1}{1 + \beta} \tag{4.3}$$

From Equation 4.3, the $\hat{C_D}$ will be decreased compared to the C_D for the simple model in Equation 4.1. The difference can be illustrated by substituting values into the expression for α and β. Consider a nodule that increases 20% in volume from the measurement on the first scan to the second scan, that is, $V_{\text{T2}} = 1.2V_{\text{T1}}$, or $N_{\text{T2}} = 1.2N_{\text{T1}}$. If the ROI includes voxels from the attached structure that correspond to 10% of the nodule volume, so that $N_{\text{att}} = 0.1N_{\text{T1}}$, then the $\hat{C_D}$ would be

$$\hat{C_D} = \frac{\alpha - 1}{1 + \beta} = \frac{1.2 - 1}{1 + 0.1} = \frac{0.2}{1.1} = 0.182$$

which is lower than the actual volume change of 20%. This implies a reduced sensitivity to nodule growth, with larger values of β resulting in smaller $\hat{C_D}$ values.

We wish to find α in the presence of β to obtain an accurate measure of nodule growth. To mitigate the impact of attached structures, a non-uniform weighting function can be used that decreases the weight of the attached structures. If we make the assumption that the attached structures will be far from

the center of the nodule, then a weighting function such as the one illustrated in Figure 4.6 can be used, which reduces the weight of voxels that are far from the nodule center. This weighting function will decrease β proportionally by some amount B, where $0 < B < 1$, but it will also decrease the value of α proportionally by A. The relative density change with weighting (C_{DW}), becomes:

$$C_{DW} = \frac{A\alpha - 1}{1 + B\beta}$$

An ideal weighting function would apply no weight ($B = 0$) to the non-nodule structures and all the weight ($A = 1$) to the voxels belonging to the nodule. This is difficult to achieve in practice, due to the uncertainty in the boundary between the nodule and non-nodule soft tissue structures, but as long as A is greater than B, applying a weighting function will result in a better estimate of nodule growth. Using the previous example with a 20% change in the nodule size, and attached structures that have a volume equal to 10% of the nodule volume, the C_{DW} is:

$$C_{DW} = \frac{A\alpha - 1}{1 + B\beta} = \frac{1.2A - 1}{1 + 0.1B}$$

If $A = 0.99$ and $B = 0.1$, then

$$C_{DW} = \frac{1.2 \cdot 0.99 - 1}{1 + 0.1 \cdot 0.1} = \frac{0.188}{1.01} = 0.186$$

which is slightly higher than the non-weighted case. Note that, if A is significantly different than 1, the C_{DW} value decreases substantially, so it is important to select a weighting function that includes as many of the voxels within the nodule as possible.

A 3D isotropic Gaussian weighting function was chosen for preliminary study. The weight of a particular voxel was solely dependent on the distance from the center of the nodule according to a Gaussian function, which was given

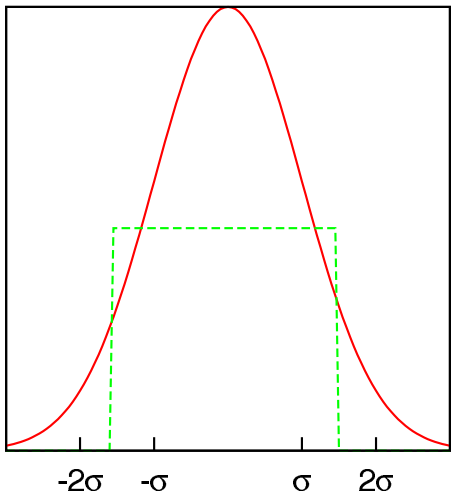


Figure 4.7: Diagram showing an idealized nodule profile in 1D (dashed line) with a Gaussian (solid line). The width of the Gaussian is selected to reduce the impact of structures far from the nodule center while still including the entire nodule.

by the following equation:

$$G(x, y, z, x_c, y_c, z_c, \sigma) =$$

$$\frac{1}{(2\pi)^{3/2}\sigma^3} \exp\left(\frac{-(x-x_c)^2 - (y-y_c)^2 - (z-z_c)^2}{2\sigma^2}\right)$$

where x_c, y_c, and z_c is the coordinate of the center of the nodule, and σ is the standard deviation of the kernel. The components at the center of the nodule have the greatest contribution to the mean density, and structures far from the center have less of a contribution. A diagram illustrating an idealized nodule intensity profile in 1D and a Gaussian function is shown in Figure 4.7.

A critical parameter of the Gaussian weighting function is the size, σ. If σ is too small, the weighting function will apply a low weight to some voxels that are part of the nodule, and as a result, it will not be very sensitive to changes in the nodule size. However, if σ is too large, non-nodule voxels will have a relatively high weight, and therefore will be included in the growth calculation. Thus, the mean density was computed according to the following equation on

both images:

$$M = \sum_x \sum_y \sum_z I(x,y,z)\, G(x,y,z,C_2,S_2)$$

Note that, unlike the mean density computed using a uniform weighting function, the Gaussian-weighted mean density is not linearly related to the volume. Te determine the relationship, the expression for the Gaussian-weighted mean density for a sphere was derived. Given a a sphere of radius r and intensity A, and an isotropic Gaussian G_σ with a standard deviation of σ, the mean density is:

$$\begin{aligned} M &= A \iiint_V G_\sigma \mathrm{d}V \\ &= A \int_{-R}^{R} \mathrm{d}z \int_{-\sqrt{R^2-z^2}}^{\sqrt{R^2-z^2}} \mathrm{d}y \int_{-\sqrt{R^2-z^2-y^2}}^{\sqrt{R^2-z^2-y^2}} \mathrm{d}x\; G_\sigma(x,y,z) \end{aligned}$$

To solve this expression, the rectangular coordinates are converted into spherical coordinates.

$$\begin{aligned} M &= A \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi \int_0^{R} \mathrm{d}r\; \frac{\sin(\phi) r^2}{(\sqrt{2\pi}\sigma)^3} e^{-\frac{r^2}{2\sigma^2}} \\ &= \frac{A}{2\pi\sigma^2} \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi\; \sin(\phi) \int_0^{R} \mathrm{d}r\; \frac{r^2}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} \\ &= \frac{A}{2\pi\sigma^2} \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi\; \sin(\phi) \left(\int_0^{R} \frac{r^2-\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} \mathrm{d}r + \sigma^2 \int_0^{R} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} \mathrm{d}r \right) \\ &= \frac{A}{2\pi\sigma^2} \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi\; \sin(\phi) \left(\left[-r\sigma^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} \right]_{r=0}^{r=R} + \frac{\sigma^2}{2} \mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}} \right) \right) \\ &= \frac{A}{2\pi\sigma^2} \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \mathrm{d}\phi\; \sin(\phi)\sigma^2 \left(-R \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{R^2}{2\sigma^2}} + \frac{\sigma^2}{2} \mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}} \right) \right) \end{aligned}$$

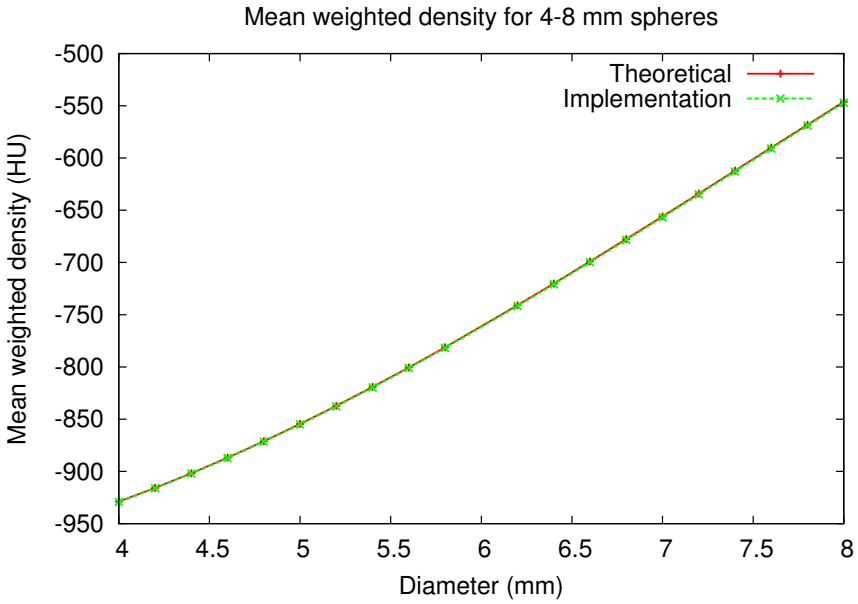


Figure 4.8: Mean weighted density, computed theoretically and according to the implemented algorithm, for simulated image data of spheres changing from 4 to 8 mm in diameter

$$
\begin{aligned}
M &= \frac{A}{2\pi}\left(-R\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{R^2}{2\sigma^2}}+\frac{\sigma^2}{2}\mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}}\right)\right)\int_0^{2\pi}\mathrm{d}\theta\ \left[-\cos(\phi)\right]_{\phi=0}^{\phi=\pi} \\
&= \frac{A}{2\pi}\left(-R\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{R^2}{2\sigma^2}}+\frac{\sigma^2}{2}\mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}}\right)\right)\int_0^{2\pi}\mathrm{d}\theta\ 2 \\
&= \frac{A}{2\pi}\left(-R\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{R^2}{2\sigma^2}}+\frac{\sigma^2}{2}\mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}}\right)\right)\left[2\theta\right]_{\theta=0}^{\theta=2\pi} \\
M &= 2A\left(-R\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{R^2}{2\sigma^2}}+\frac{\sigma^2}{2}\mathrm{erf}\left(\frac{\mathrm{R}}{\sqrt{2}\mathrm{œ}}\right)\right) \qquad (4.4)
\end{aligned}
$$

Since the expression for the mean weighted density M is comprised of the error function(erf), the expression has no closed-form solution, though numerical approximations exist. To validate the developed code, the program was applied to a series of images of spheres 4 to 8 mm in diameter. The images were generated by a VisionX program, `vgsphere`, that generates a gray-level image of a sphere taking into account partial voxels along the boundary. The generated spheres

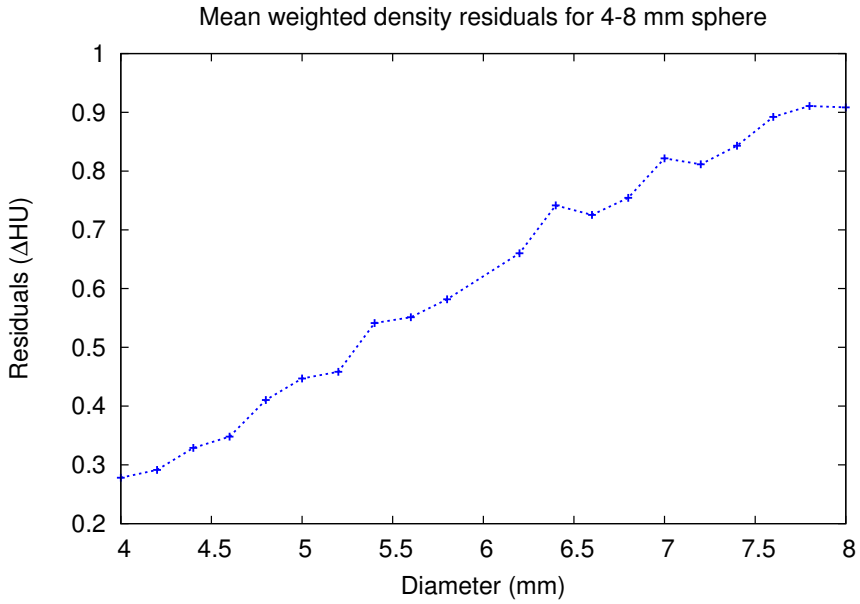


Figure 4.9: Residuals for theoretical - implemented algorithm for simulated spheres of size 4 to 8 mm in diameter

had an intensity of 0HU. The 8 mm sphere was used to set the σ for the Gaussian. The mean density computed from Equation 4.4 using GNU Octave [67] was plotted with the mean density computed from the implemented algorithm in Figure 4.8, with the residuals shown in Figure 4.9. The residuals were less than one HU in all cases.

The next step was to assess the relationship between the measurements from the mean weighted density and the nodule volume. Since the expression for growth index and relative size change both involve the term $\frac{S_2}{S_1}$, where S indicates the size determined by a given method, this was the quantity considered in the analysis. The typical range of interest for measuring nodule growth is up to a doubling in nodule volume. For the case of a nodule 4 mm in diameter, the volume is 33.5 mm^3, so doubling the volume to 67.0 mm^3 results in a nodule diameter of 5.0 mm, and halving the volume to 16.8 mm^3 results in a nodule diameter of 3.2 mm. Using the relative size computed from the mean density

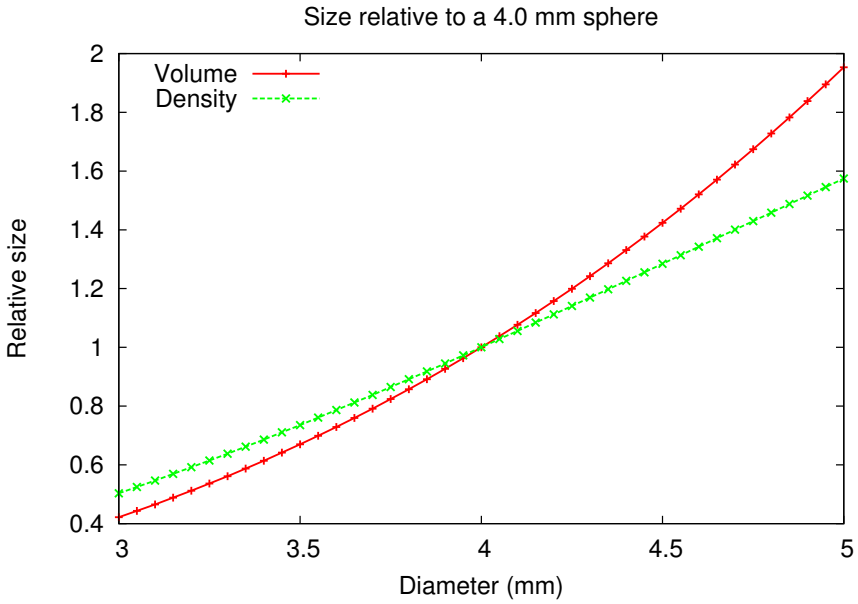


Figure 4.10: Comparison of relative size according to mean weighted density and volume. Since the relative size computed using the mean weighted density differs from the curve for the relative size computed from volume, a transformation function was necessary to make the values directly comparable.

directly (relative to a 4.0 mm diameter sphere) for the range of nodule diameters of 3.0 mm to 5.0 mm results in the relationship shown in Figure 4.10. Nodules with diameters less than 4.0 mm are considered to have decreased in size, while larger than 4.0 mm have increased in size. The relative size computed from the mean density varies differently with the nodule diameter than volume. Although this does not have an impact on diagnostic performance, this makes it difficult to directly compare the density change measurements with volume.

An expression for the relationship between the mean density change and the volume change can be numerically estimated from the measurements in Figure 4.10. Based on the plot, the data approximated a quadratic function in the interval of interest:

$$\mathrm{RV} = a \cdot \mathrm{RD}^2 + b \cdot \mathrm{RD} + c \tag{4.5}$$

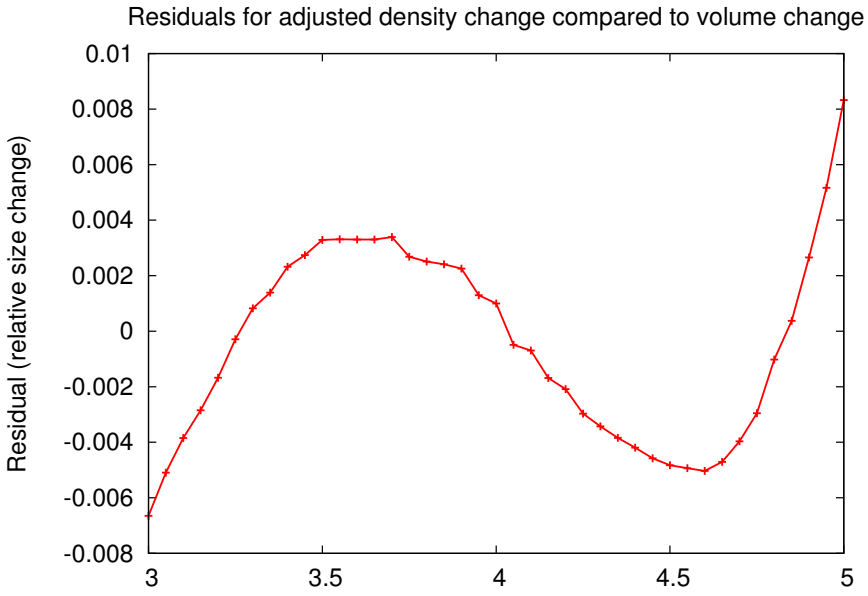


Figure 4.11: Residuals of the density change compensated according to Equation 4.5 compared to volume change.

where RV $= \frac{V_2}{V_1}$ and RD $= \frac{M_2}{M_1}$ using the non-linear least squares algorithm in R [68]. The coefficients were estimated to be $a = 0.465$, $b = 0.449$, and $c = 0.085$. An expression without the intercept term c was also tested, but was found to have a fit that was worse. Other, more complex expressions were tested, including higher powers and exponential functions, and none resulted in statistically significant lower error. The residuals of the new, adjusted mean density change, are shown in Figure 4.11. The residuals are very small (less than 1 percent) in the interval of interest.

4.2 Growth index from density (GI_D) method

Based on the nodule model described in Section 4.1, a method was developed to measure nodule growth index from the change in density (GI_D). The major steps of the GI_D method are shown in Figure 4.12. First, preprocessing of the CT

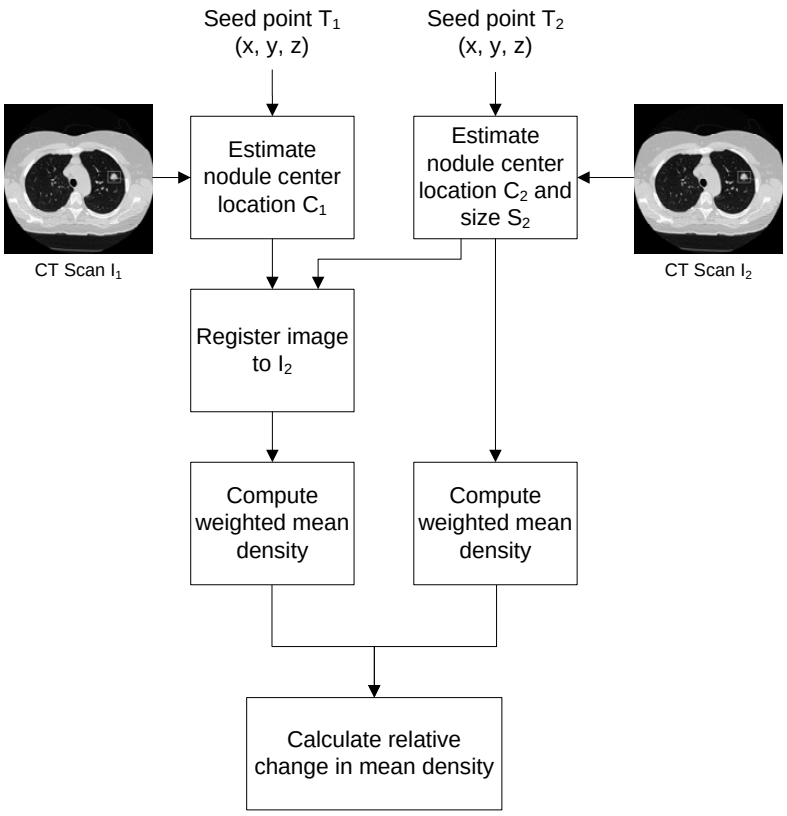
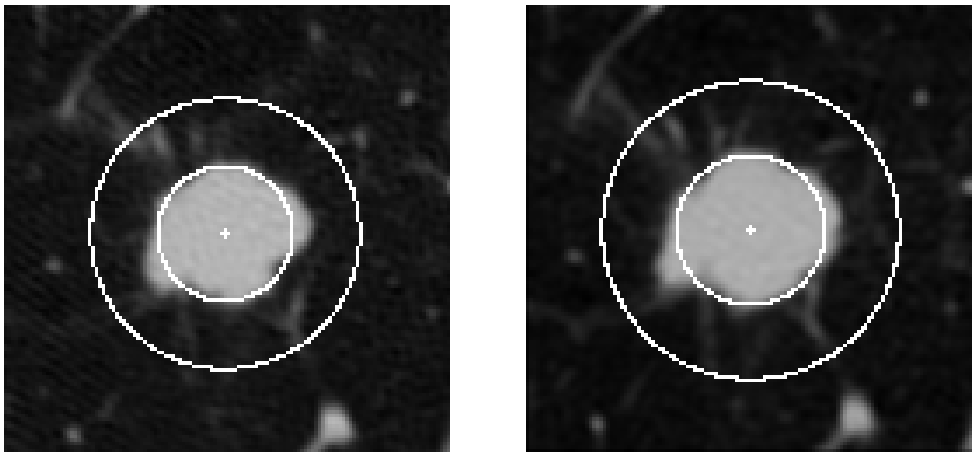


Figure 4.12: Overview of growth index from density measurement algorithm

scans was performed to select a region of interest around the nodule and generate an isotropic image. Next, the center and size of the nodule are estimated from a manually specified seed point using an iterative optimization approach. In the third step, the regions surrounding the nodule on each CT scan were registered. In the final step, a weighting function is applied to the region and the mean density computed; this mean density is used as a surrogate for volume in the computations for growth rate. The first three steps relied on previously published work [23] described in Section 1.6. The final step is novel to this method. The method is evaluated for variability and accuracy of diagnosis using several datasets.

4.2.1 Region of interest preprocessing

In the first step, regions of interest were extracted from the full CT scan and prepared for later steps of the algorithm. The input to the algorithm was a pair of CT scans, I_1 and I_2, containing the nodule, and seed points located within the nodule on both scans. Based on this, for each scan, a region of fixed size was extracted from the original CT scan around each seed point. These regions were resampled into isotropic space using trilinear interpolation, and these images, I_{R_1} and I_{R_2}, along with the coordinates of the seed points in the resampled space were provided to the next stage of the algorithm. The primary reason for resampling the image was to allow for subvoxel precision for locating and sizing the nodule. However, resampling the image to a higher resolution has the drawbacks of increasing the image size and computation time. An additional step of juxtapleural detection was performed to aid in the next step.

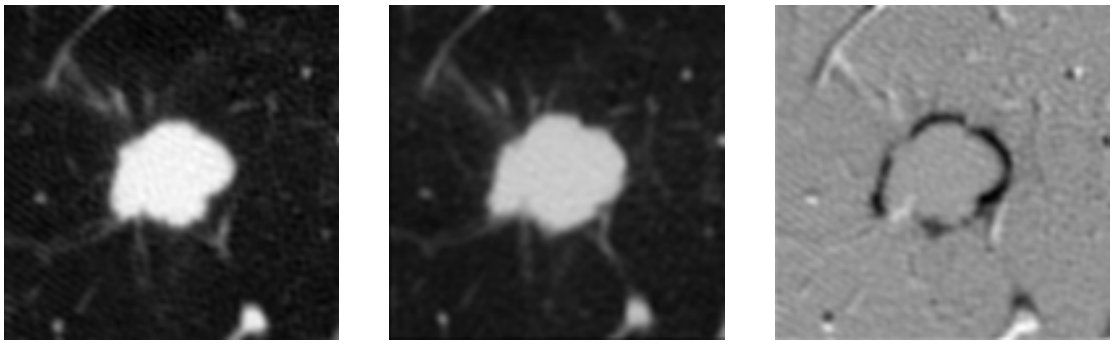


(a) (b)

Figure 4.13: Estimate of nodule location and size for a) scan at T1 and b) scan at T2. Only the central slice of the region is shown. The location of the nodule is at the center of the circle; the inner circle indicates the size estimate for the nodule.

4.2.2 Estimation of nodule center and size

Given the resampled regions and seed points, estimates of the nodule center, C_1 and C_2, and size, S_1 and S_2, on both regions were determined using the Reeves et al. algorithm described in Section 1.6, but with one modification – the estimation was performed on the resampled isotropic image to achieve superior location resolution. An example of the result of the algorithm on the central slice of the nodule on the first and second scans is shown in Figure 4.13. An initial estimate may, optionally, be provided by the user to the algorithm; this estimate will be used as an initial condition and the same process described above is used to determine the nodule size. Note that the computed nodule center and size from this method are only estimates used to determine the appropriate location and kernel size for later steps in the algorithm. The region from the second scan was reduced in size based on the estimated nodule size; the region of the first scan is not altered in this step.



(a) (b) (c)

Figure 4.14: Result of region registration, central slice of a) nodule region on first scan registered to b) second scan and c) difference image between the registered image and the second scan. Gray indicates no difference.

4.2.3 Nodule region registration

Since the mean weighted density will be computed over regions of interest, the regions must be comparable from one scan to another. To accomplish this, the regions were registered – the I_{R_1} image from Section 4.2.1 was registered to the second scan, I_{R_2} using an intensity-based three-dimensional rigid body registration algorithm [23]. In brief, the algorithm used a Gaussian-weighted mean-square-difference matching metric and conducted a search over all parameters to minimize the metric using Powell's method. For this method, the initial translation parameter was derived from the computed difference in nodule centers, $C_2 - C_1$, and the size of the Gaussian weighting function was derived from the estimated size of the nodule on the second scan (S_2). In Figure 4.14, the nodule on the first scan has been registered to the nodule on the second scan. The registered image I_{RR_1} is shown on the left with the difference between the registered image and the nodule on the second scan on the right ($I_{R_2} - I_{RR_1}$). Note that this step does not alter the resampled image from the second CT scan.

4.2.4 Density-based growth estimation

Algorithm 1: Density-based growth index measurement

input : Resampled, registered image I_{RR_1} and resampled image I_{R_2}
input : Location at T1 C_1 and T2 C_2
input : Size at T2 S_2
input : Days between T1 and T2, ΔT

$\sigma \leftarrow \omega S_2$
forall the $W_i \in \{I_{RR_1}, I_{R_2}\}$ **do**
  forall the $w \leftarrow W_i$ **do**
    $v_i \leftarrow$ `WeightingFunction`$(w, x, y, z, x_c, y_c, z_c, \sigma)$
    $M_i \leftarrow M + v$
  $M_i \leftarrow M_i / |W_i|$
$\mathrm{GI_D} \leftarrow$`CalcuateGI`$(M_1, M_2, \Delta T)$

Once the two registered images are obtained, the next step is to compute the growth index for the nodule according to Algorithm 1.

4.2.4.1 Weighting function

The size of the function, σ, is proportional to the size of the nodule on the second scan (S_2) by a constant ω, depending on the weighting function. Two different weighting functions were tested: a uniform sphere and a 3D Gaussian. The uniform sphere was centered at (x_c, y_c, z_c) with a radius of σ. The size of the sphere, σ, was set to the size of the nodule on the second scan ($\omega = 1$) to ensure all the majority of the voxels within the nodule are considered. Additional details about both weighting functions are provided in Section 4.1.

4.2.4.2 Calculate $\mathrm{GI_D}$

The $\mathrm{GI_D}$ computed using the uniform sphere weighting function, $\mathrm{GI_{UD}}$, was computed according to the following equation:

$$GI_{UD} = 100 \left[(M_2/M_1)^{30.4375/\Delta T} - 1 \right] \tag{4.6}$$

where M_1 and M_2 are the estimators measured by density-based growth method method on the first and second scans, Δt is the interval in days between scans.

The 3D Gaussian function was introduced in Section 4.1.The Gaussian weighting function was centered at (x_c, y_c, z_c) and the size of the weighting function, σ_g, was set to 66% of the estimated nodule radius ($\omega = 0.66$) and truncated at 3σ, so that voxels located further than twice the estimated nodule radius would not be considered when calculating the mean density (weight of 0). The growth index from weighted density (GI_{WD}) was computed using Equation 4.5:

$$GI_{WD} = 100 \left[\left((a(M_2/M_1)^2 + b(M_2/M_1) + c \right)^{30.4375/\Delta t} - 1 \right] \tag{4.7}$$

using the same parameters GI_{UD}, with additional coefficients $a = 0.465$, $b = 0.449$, and $c = 0.085$, as derived in Section 4.1.

4.3 GI_D evaluation

To evaluate the GI_D methods, two performance metrics were used: interscan variability and diagnostic performance. Interscan variability indicates the expected range of measurements of the same nodule on a different scan, assuming no change in size. In the ideal case, a method would measure no change in nodule size. Higher variability increases the confidence interval on the growth measurement, which results in greater uncertainty of the diagnosis. The second metric measures how well the growth indices estimated by the GI_D method discriminate between benign and malignant nodules.

The GI_D method was evaluated using both a uniform spherical weighting function (GI_{UD}) and a 3D Gaussian weighting function (GI_{WD}). Measurements

Table 4.1: Parameters of datasets

Dataset	# nodules	Slice Thickness (mm)	kVp	Current (mA)
Zero-change	20	1.25 - 5.0	120 - 140	40 - 250
Stable	38	1.0, 1.25	120 - 140	80 - 300
Malignant	19	1.0 - 5.0	120 - 140	40 - 300
Complex	4	1.0 - 5.0	120, 140	40 - 300

Table 4.2: Nodule size information

Dataset	# nodules	Mean, SD Size (mm)	Median, range of interval (days)
Zero-change	20	12.69, 3.63	0
Stable	38	6.91, 3.16	393.5, 91 - 1918
Malignant	19	6.68, 3.03	165, 90 - 756
Complex	4	17.61, 3.81	50, 29 - 98

from a volumetric method were computed compared to the GI_D method to assess correlation.

4.3.1 GI_D interscan variability

The interscan variablities of the GI_D and volumetric (GI_V) methods were evaluated by measuring the percentage size change (PSC), defined as C_{UD} and C_{WD} for the GI_{UD} and GI_{WD} methods respectively and by percentage volume change for the GI_V method, on 20 zero-change cases. The interscan variability was defined to be represented as the 95% interval of measurements given by $(\mu - 1.96\sigma,\ \mu + 1.96\sigma)$. Each of the 20 cases had a single nodule imaged on scans several minutes apart during the preliminary stages of a biopsy. Consequently, the actual volume change of the nodules between these scans was zero. Ten cases had two scans of the same slice thickness (1.25 mm), while ten cases had two scans of differ-

ent slice thicknesses – in seven of these cases, one scan had a slice thickness of 1.25 mm and the other was 2.5 mm, and in three cases, one scan was 1.25 mm while the other was 5.0 mm. The twenty cases were selected from the Prevent Cancer Research Foundation database of pulmonary nodules[1] according to the following criteria:

- Solid consistency, as determined by a radiologist
- Successful volumetric segmentation
- Two scans through the entire nodule

The nodules were imaged using either a GE LightSpeed QX/i or LightSpeed Ultra scanner, and the nodules were all less than 20 mm in size, as measured by a volumetric measurement method. Additional parameters of the dataset are provided in Table 4.1 and Table 4.2.

4.3.1.1 Interscan variability results

The results of the interscan variability experiment for the GI_{UD}, GI_{WD}, and the volumetric method are presented in Table 4.3 for the full dataset (20 nodules) and the same slice and mixed sliced subsets (10 nodules each). In two cases, the nodule location and size estimation step failed; for these cases, the methods were provided a user-estimated radius in addition to the seed point.

4.3.1.2 Interscan variability discussion

The interscan variability experiment measured the variation in size measurement for nodules with no change for different subsets of nodules. For both

[1] Available from http://www.via.cornell.edu/databases/crpf.html

Table 4.3: Interscan variability reported as percentage size change (PSC) for GI_D method with uniform (GI_{UD}) and 3D Gaussian weighting (GI_{WD}) functions and a volumetric (GI_V) method

Dataset	Method	Mean PSC (%)	SD PSC (%)	95% interval of PSC
Same slice	GI_V	-3.3	11.2	-25.2, 18.6
	GI_{UD}	-0.2	6.4	-12.7, 12.4
	GI_{WD}	0.6	6.0	-11.0, 12.3
Mixed slice	GI_V	-21.0	25.6	-71.2, 29.2
	GI_{UD}	-6.6	9.1	-24.5, 11.2
	GI_{WD}	-5.0	10.4	-25.3, 15.5
Full	GI_V	-12.2	21.6	-54.6, 30.3
	GI_{UD}	-3.4	8.5	-20.1, 13.3
	GI_{WD}	-2.2	8.9	-19.7, 15.3

methods, the lowest interscan variability was observed on the subset of nodules imaged on scans of the same slice thickness. This is expected, since using scans of the same slice thickness removes a source of variation (change in slice thickness). On this subset, the volumetric method had similar interscan variability to other previously published studies summarized in Section 1.4.4, which is approximately -20% to 20%, but the GI_D methods had much lower interscan variability using both weighting functions. These results reinforce the importance of ensuring that longitudinal scans are taken using the same slice thickness.

On all the sets of nodules, the GI_{UD} method had much lower interscan variability than the GI_V method. The percentage size change (PSC) measurements were significantly different between the two methods ($p = 0.019$, Wilcox signed-rank test). The largest difference in interscan variability for the GI_{UD} and GI_V methods was on the subset of nodules imaged on scans of different slice thicknesses where the GI_{UD} method had nearly 3 times less variability than the GI_V

method. The decreased variability suggests that the GI_{UD} method will allow for less uncertainty in the growth index measurement.

The interscan variability for the GI_{WD} was similar to the GI_{UD} method; the PSC measurements did not differ significantly ($p = 0.294$, Wilcox signed-rank test) between the GI_{WD} and GI_{UD} methods on this dataset. The similarity between the GI_{WD} and GI_{UD} methods was not surprising – since these are zero-change nodules, the ROIs should remain the same from one scan to the other and we would expect the GI_{WD} and GI_{UD} methods to give similar results. The variability was slightly lower on the same slice thickness subset of nodules compared to the mixed slice thickness subset. This may be due to the additional variation in the ROI appearance in the mixed slice thickness subset.

One of the assumptions of the GI_D method is that a nodule will have an approximately spherical shape. If a nodule deviates significantly from a sphere, such as a lengthy, flat nodule, the ROI will include either include a substantial number of non-nodule voxels or exclude a number of nodule voxels, thereby altering the apparent size of the nodule. However, in this dataset, all the nodules were approximately spherical.

Despite these limitations, on the datasets evaluated in this study, the GI_D methods provided more consistent measurements than the volumetric method. However, for a nodule growth measurement to be effective, it must not only be consistent, but it must also be sensitive to growth. The next section evaluates the performance of the method for measuring nodule growth.

4.3.2 GI_D Diagnostic performance

The motivation for measuring pulmonary nodule growth index is to diagnose malignant nodules. This experiment evaluated how well the growth rates from

the GI_D and volumetric methods (GI_V) predicted malignancy by using datasets of benign and malignant nodules to establish a growth rate threshold for malignancy. There were two steps to the evaluation:

1. Use the benign dataset to establish a threshold on the growth index, above which nodules would be considered malignant
2. Apply this threshold to the malignant dataset to measure the diagnostic performance

To establish the growth index threshold, the upper value of the 95% interval of the growth indices of the benign nodules were computed. The upper 95% value, T, was computed using the following equation:

$$T = \mu + 1.96 \cdot \sigma$$

where μ is the mean growth index for the benign nodules and σ is the standard deviation of the benign nodules growth indices. This threshold was used to classify nodules from both the the malignant and benign datasets, and the accuracy was reported for each dataset.

Another method to evaluate the diagnostic performance is to produce a receiver operating characteristic (ROC) curve, which plots the sensitivity (number of malignant nodules correctly classified) against 1-specificity (number of benign nodules correctly classified) for varying thresholds.

The benign nodules were confirmed by either biopsy or the absence of clinical change in a two year time period, while the malignant nodules were confirmed by biopsy or resection. All nodules were selected from the Weill Cornell Medical College database according to the same criteria described in Section 4.3.1. There were 38 benign nodules, and all of the nodules but one had scans of the same slice thickness. Scans were acquired with either a GE LightSpeed Ultra,

LightSpeed Pro 16, LightSpeed VCT, HighSpeed CT/i, or Genesis scanner. The dataset had 19 malignant nodules, which were selected to have an interval of at least 90 days and a volume change of at least 30%, as measured by the volumetric method described in Section 1.6. These criteria were selected to ensure that the volume change was greater than the measurement uncertainty measured by previous studies. An additional four malignant nodules with a complex appearance were included with scans taken at intervals ranging from 29 to 98 days, for a total of 23 malignant nodules. The scans for the malignant nodules were acquired using the following GE scanners: LightSpeed Ultra, LightSpeed QX/i, HighSpeed CT/i, and Genesis. Additional parameters for both the malignant and benign nodules are provided in Table 4.1 and Table 4.2.

For the volumetric method, the growth index was computed using the volume instead of the density:

$$\mathrm{GI_V} = 100\left[(V_2/V_1)^{30.4375/\Delta T} - 1\right]$$

where V_1 and V_2 are the volumes computed by the volumetric method on the first and second scans and ΔT is the interval in days between scans.

4.3.2.1 Diagnostic performance results

The diagnostic performance was determined by applying a threshold to the growth index for each nodule, and determining whether the nodule was classified correctly. The GI range for the benign and malignant datasets are shown in Table 4.4. Using the thresholds given in Table 4.5 computed from the benign nodule dataset, the $\mathrm{GI_D}$ and volumetric methods classified the nodules according to their GI. The diagnostic performance for the $\mathrm{GI_D}$ and volumetric methods

Table 4.4: Growth index statistics for both methods on the stable and malignant datasets

Dataset	Method	GI Range	Median GI
Benign	GI_V	-6.2 – 6.4	0.10
	GI_{UD}	-2.9 – 1.4	0.043
	GI_{WD}	-3.3 – 1.7	0.17
Malignant	GI_V	3.0 – 45.5	16.7
	GI_{UD}	1.7 – 12.6	5.8
	GI_{WD}	1.9 – 18.1	6.9

Table 4.5: Malignancy thresholds determined from stable nodules.

Method	Mean GI	SD GI	Threshold for malignancy (%/month)
GI_V	0.07	2.45	4.87
GI_{UD}	-0.07	0.83	1.56
GI_{WD}	-0.03	1.00	1.92

are provided in Table 4.6. The GI malignancy threshold was set from the upper limit of the 95% interval for the benign nodules. Note that the performance for the benign nodules is optimistic, since these nodules were used to establish the threshold.

The ROC curve is shown in Figure 4.15 for the classification performance for the GI_D and volumetric methods for the entire dataset of benign and malig-

Table 4.6: Nodules correctly classified based on GI for GI_{UD}, GI_{WD}, and GI_V methods

Method	Benign	Malignant
GI_V	92.1% (35/38)	73.9% (17/23)
GI_{UD}	100% (38/38)	95.7% (22/23)
GI_{WD}	100% (38/38)	95.7% (22/23)

nant nodules. The volumetric method failed to segment four nodules, so for the purpose of making the ROC curve, the GI values for these nodules was set as 0%/month. The area under the ROC curve (AUC) was 1.00 for GI_{UD} , 0.997 for GI_{WD}, and 0.897 for GI_V. The AUC for GI_{UD} was higher than for GI_{WD}, despite both methods having identical performance in 4.6; this was due to the method by which the classification threshold was selected. The lowest malignant GI_{UD} was 1.54 %/month, while the highest benign GI_{UD} was 1.39 %/month and thus, no confusion between the malignant and benign nodules.

For these nodules, the differences in growth index values for the benign nodules were not statistically significant ($p = 1.00$, Wilcox signed-rank test for GI_{UD} and GI_V , $p = 0.84$ for the GI_{WD} and volumetric methods) between the GI_D and volumetric methods, but the growth indexes for the malignant nodules were statistically significantly different ($p < 0.01$, Wilcox signed-rank test). The differences in GI for the benign nodules between the GI_{UD} and GI_{WD} methods were not significantly different ($p = 0.71$), but there was a significant difference ($p < 0.01$) for the malignant nodules. The GI_{WD} method incorrectly classified one of the 19 malignant nodules, but correctly classified all four of the complex malignant nodules, while the GI_{UD} method correctly classified all 19 malignant nodules, but only half (2/4) of the complex malignant nodules.

The GI_{UD} values were plotted against the GI_V values for the benign and malignant nodules in Figures 4.16 and 4.17. For the plot of GI_{UD} versus GI_V, the parameters of the best fit line, $y = a \cdot x + b$, were $a = 0.318$ and $b = 0.151$, with a coefficient of determination $r^2 = 0.905$, while for the GI_{WD}, the coefficients were $a = 0.280$ and $b = 0.142$, with a coefficient of determination $r^2 = 0.880$.

Finally, the GI_{WD} values were plotted against the GI_{UD} values in Figure 4.18.

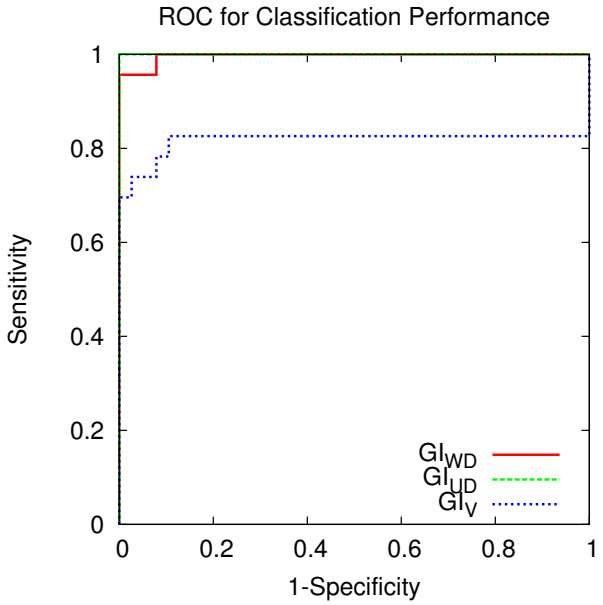


Figure 4.15: ROC curves showing diagnostic performance for GI_{WD}, GI_{UD}, and GI_V

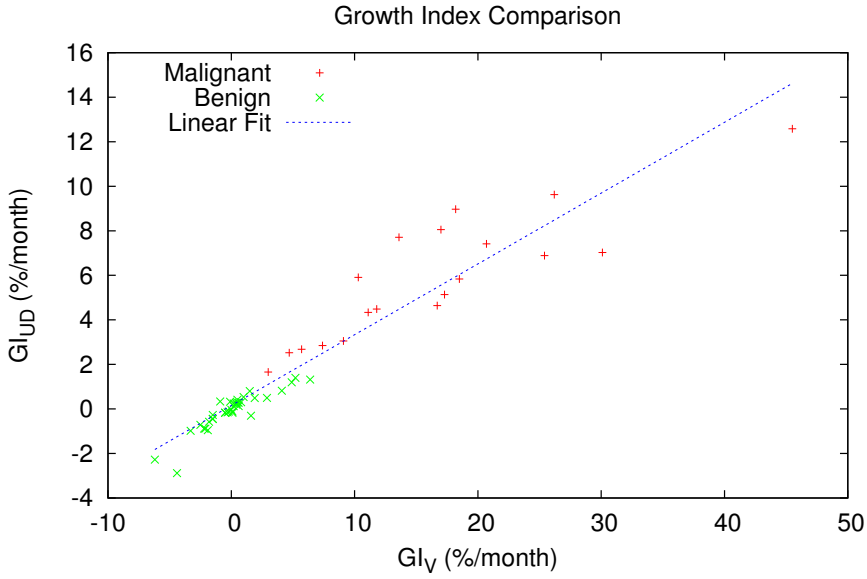


Figure 4.16: Plot of GI_{UD} versus the GI_V indicating a linear relationship

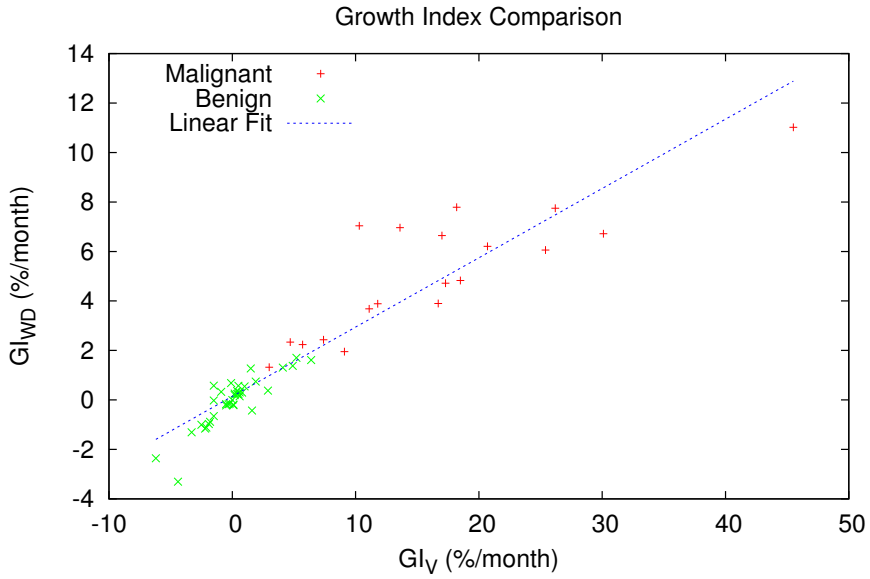


Figure 4.17: Plot of GI_{WD} versus GI_V values indicating a linear relationship with a similar appearance to 4.16

The parameters of the best fit line were $a = 0.884$, $b = 0.002$, and $r^2 = 0.981$.

4.3.2.2 Diagnostic performance discussion

The main purpose for measuring pulmonary nodule growth index is to aid in establishing the malignancy status of a suspicious nodule. The diagnostic performance of the GI_D method, with either the uniform weighting or 3D Gaussian weighting, was significantly better than the volumetric method, especially for the malignant nodules. Compared to the volumetric method, five additional malignant nodules were correctly classified by GI_{WD}, and four malignant nodules by the GI_{UD} method. Three additional benign nodules were correctly classified by both density methods. This improvement in performance was especially impressive considering the density methods were much simpler than the volumetric method. All four complex malignant failed to be segmented correctly by the volumetric method, which led to lower performance on the malignant nod-

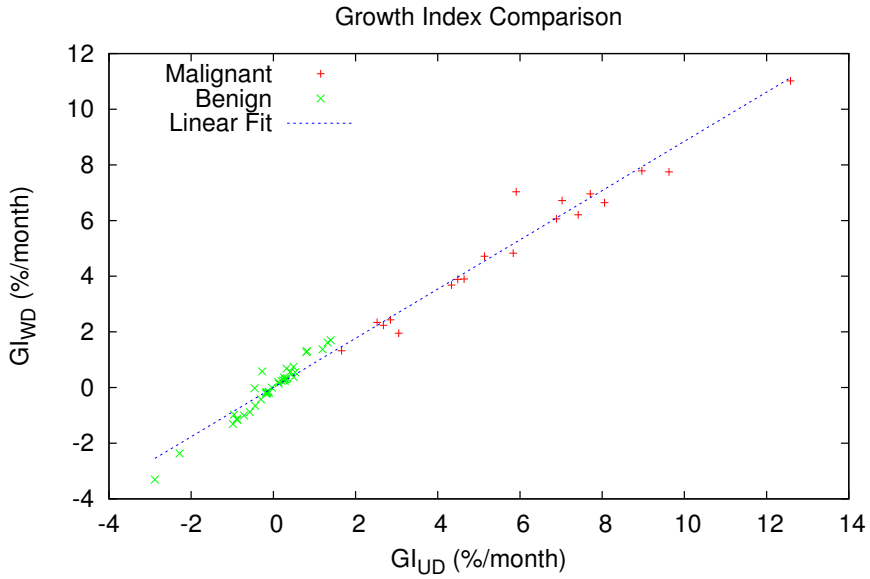
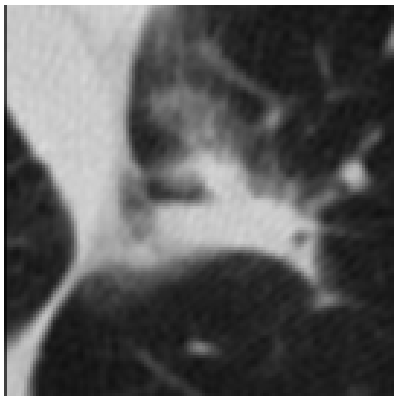


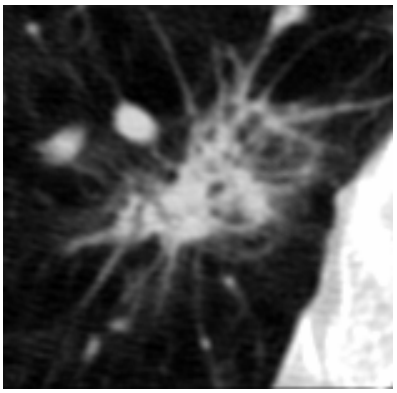
Figure 4.18: Plot of GI_{WD} versus GI_{UD} values that show a close relationship between the two methods, as indicated by the linear relationship with a slope near 1

ules than the density methods. Two of the four complex nodules (50%) were correctly classified by the GI_{UD} method, while all four were correctly classified by the GI_{WD} method; a central slice of these nodules is shown in Figure 4.19.

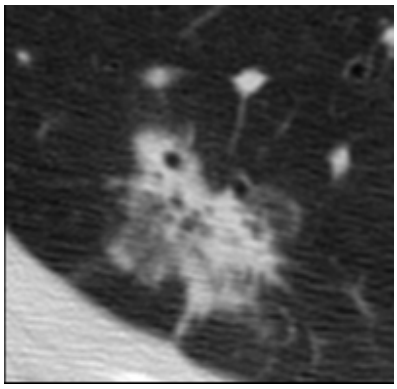
All the methods had GI thresholds that were consistent with stability according to other studies [39], which consider nodules with a GI less than 5.4%/month to be benign, though the GI_{UD} method had a much lower threshold. The GI_{UD} method had much smaller ranges for the GI, compared to the volumetric method, for both the malignant and benign nodules, and the median GI for the malignant nodules was closer to zero for the GI_{UD} method. This might suggest that the GI_{UD} and volumetric methods are not well correlated. Plotting the GI_{UD} results against the GI_V results showed a linear relationship, but the best fit linear model differed from the identity line, which indicates some differences in the measurements between the two methods. This is likely attributable to the fact



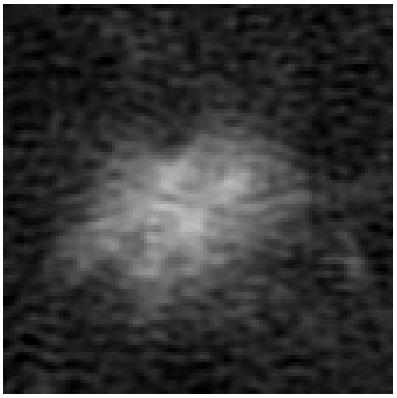
(a)



(b)



(c)

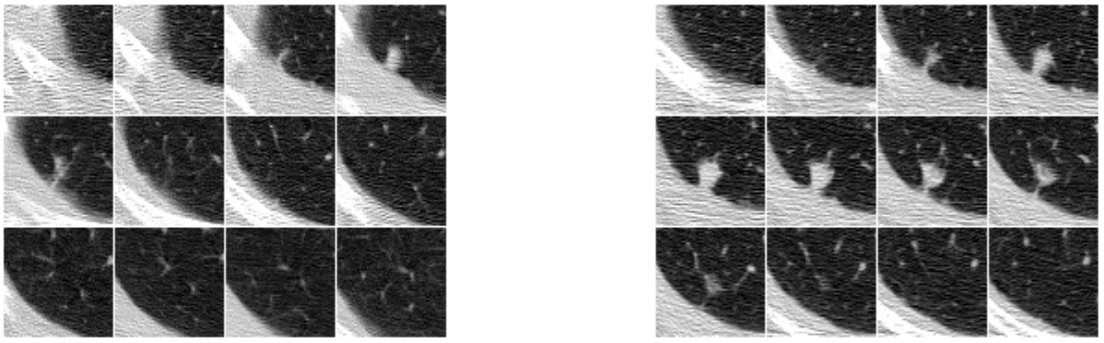


(d)

Figure 4.19: Central slice of three nodules which were not successfully segmented by the volumetric method. All these nodules are malignant. Images are not to scale.

that the GI_{UD} method includes all voxels belonging to any soft tissue in calculating the nodule size, while the GI_V method only includes voxels that it considers to be part of the nodule. To address this problem, the next section introduces a weighting function that was modified to decrease the weight of structures further from the nodule center.

The GI_{WD} method was developed to address nodules with a complex appearance where the boundary between the nodule and lung parenchyma is uncertain, and for these cases, there was a substantial benefit to the GI_{WD} method over the GI_{UD} method. The GI_{WD} method correctly classified all the complex malignant nodules, while the GI_{UD} method only correctly classified half of the nodules. This was responsible for a slight improvement in the AUC, from 0.95 for the GI_{UD} method to 0.99 for the GI_{WD} method. However, despite correctly classifying more of the complex nodules, this performance came at the expense of correct classification on the easier malignant nodules, where the GI_{WD} misclassified a nodule. A montage of slices through the nodule is shown in Figure 4.20. This nodule, the slowest growing malignant nodule in the dataset, has a simple appearance, except for the attachment to the pleural surface. Since the size of the nodule was estimated well, the GI_{UD} method was able to avoid considering voxels in the pleural surface in its density calculation. In contrast, the GI_{WD} method would have considered some of the voxels in the pleural surface in its density measurement, possibly a substantial amount relative to the size of the nodule on the first scan, thereby reducing the apparent growth. The GI_{WD} for this nodule was 1.32%/month, which is considerably lower than the threshold of 1.92%/month for malignancy. The GI_{UD} for the nodule was 1.65%/month, which was slightly higher, but not drastically different. The GI_{WD} method still classified more malignant nodules correctly than the volu-



(a) (b)

Figure 4.20: Several slices through the malignant nodule misclassified by the GI_{WD} method at a) Time 1 and b) Time 2. The interval between the scans was 756 days.

metric method which only correctly classified 89.5% of the malignant nodules.

The GI_{WD} and GI_{UD} values for the benign nodules were much smaller than the GI_V values. The GI_{WD} method had larger ranges of values than the GI_{UD} method, which is probably due to additional voxels which were included by the GI_{WD} method that were not part of the nodule, even though they were likely assigned a small weight. The differences in the GI values for the malignant nodules between the GI_{WD} and GI_{UD} methods were statistically significant, but not the differences in GI for the benign nodules.

The relationship between the GI_{WD} and GI_{UD} measurements were nearly linear, with a high coefficient of determination. The intercept of the best fit line was 0, indicating that the GI_{WD} was not biased against the GI_{UD} method, but the slope differed slightly from 1, which reinforces the fact that there are differences between the methods. In comparing the GI_{WD} method to the volumetric method, there was slightly less correlation between the GI_{WD} and volumetric methods ($r^2 = 0.880$) than the GI_{UD} and volumetric methods ($r^2 = 0.905$), though given the size of the datasets used in the experiment, the differences

are likely not significant.

The GI_{WD} method has many of the same limitations as described for the GI_{UD} method, but unique to the GI_{WD} method is the choice of the 3D Gaussian weighting function. For some nodules with very abrupt transitions from the nodule to the lung parenchyma and surrounding structures, applying the weighting function reduces sensitivity by including non-nodule voxels from other nearby structures, whereas the GI_{UD} method, which has a sharp cutoff at the nodule boundary, would only consider voxels belonging to the nodule. Different weighting functions, such as those with a much more abrupt transition region, may improve the performance of the GI_{WD} method without compromising its sensitivity.

4.4 Density-based growth measurement summary

The concept of measuring pulmonary nodule growth from the density change in a region is an alternative to the current and only approach of volume measurement. This density approach removes the need for explicit segmentation, which is a problem of the volumetric approach for complex nodules. The success of both the GI_{UD} and GI_{WD} methods for diagnosing pulmonary nodules is very promising compared to volumetric segmentation-based techniques. The GI_{UD} method, which uses a uniform weight for all voxels in the region of interest, performed very well, despite the presence of other soft tissue structures in the ROI. The GI_{WD} method, which applies a Gaussian weight to the ROI, classified one additional complex nodules correct compared to the GI_{UD} method, but there were nodules that could not be correctly distinguished using the growth index values provided by the GI_{WD} method.

CHAPTER 5

PILOT STUDY TO MEASURE PULMONARY NODULE GROWTH RATE IN MOUSE MODELS IMAGED WITH MICRO-CT

To aid in the quantitative evaluation of disease progression and treatment response in mouse models, a pilot study was performed to determine the feasibility of evaluating pulmonary nodule growth from micro-CT scans with an automated method. Mouse models, mice which are modified to be predisposed to developing cancer, are often used in early stage cancer research, due to their small size, ease of breeding, physiological and molecular similarities to humans, and entirely sequenced genome [69]. With the introduction of non-invasive imaging modalities, such as micro-CT, both disease progression and treatment response can be monitored in longitudinal studies, where the same animal is analyzed over a period of time. Current methods of measurement from these imaging modalities are crude and require extensive manual intervention. Not only does this require more time and effort from researchers, but it also introduces variation into the measurement; these are the same issues that automated algorithms for measurement of nodules in human lung cancer patients are designed to mitigate.

In this pilot study, the automated volumetric segmentation method described in Section 1.6 was adapted to measuring pulmonary nodules in mice imaged on micro-CT. The method was modified to address both the difference in scale between the human and mouse nodules as well as differences in the quality of the micro-CT scan. Mice are roughly 10 times smaller than humans, with a human CT scan covering approximately 30 cm compared to about 3 cm for a micro-CT scan. The micro-CT scan has additional noise and, for the particular scanner used in this study, a lack of calibration. Six nodules in four lung nodule-bearing

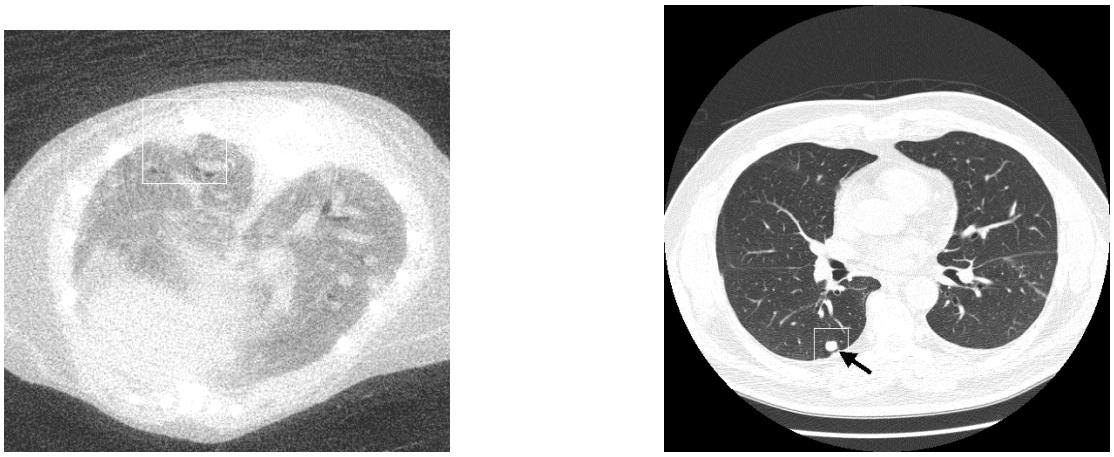
mice were sequentially scanned over a span of several months and their volumes measured by both the automated method and manual uni-dimensional measurement. Rather than use the manual measurements as the ground truth, we assumed the growth of the nodules followed the exponential model and evaluated the methods based on their fit to the model. We hypothesized that the growth rate measured by the automated algorithm would better match the exponential model than manual measurements.

5.1 Pulmonary nodules in mouse models imaged with micro-CT

Pulmonary nodules in mouse models have a similar opaque appearance on micro-CT as nodules in a human scan. Examples of a a central slice through a murine and human pulmonary nodule on CT are shown in Figure 5.1. As with a human CT scan, the micro-CT scanner has a rotating X-ray source, and the value of a voxel in the scan is proportional to the radiodensity at that location. The resulting 3D image is visualized as a series of slices. While the underlying principles are the same in micro-CT scanners, there are several differences that arise from the much smaller scale in micro-CT.

5.1.1 Differences between murine micro-CT and human CT

The micro-CT scans in this study have a resolution of $50\,\mu m \times 50\,\mu m \times 50\,\mu m$ compared to a typical resolution of $0.6\,\text{mm} \times 0.6\,\text{mm} \times 1.0\,\text{mm}$ for human CT scans. The images in Figure 5.1 show a marked increase in image noise in the micro-CT scan compared to the human CT scan and a decrease in the contrast between the parenchyma and soft tissue. In addition to these differences, there



(a) (b)

Figure 5.1: A central slice through a nodule on a) a micro-CT scan of a mouse and b) a CT scan of a human patient. Note that the images are not to the same scale.

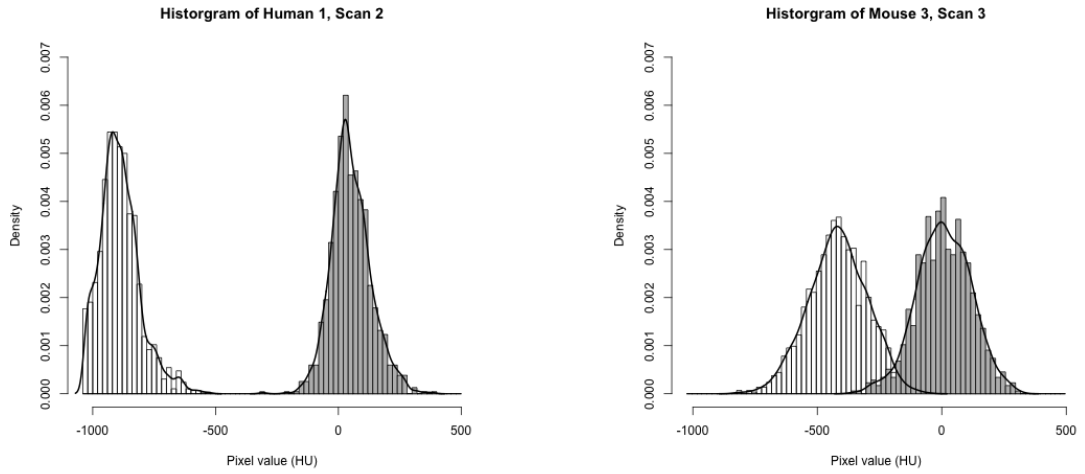
are also challenges in obtaining scans of a live mouse with the micro-CT scanner which are not a problem with modern CT scanners.

The resolution in an X-ray-based scanner is dependent on the focal spot size of the X-ray source and several parameters of the X-ray detector, including pixel size, dynamic range, and noise characteristics [70]. In order to image at the small resolutions required in micro-CT, X-ray sources with small focal spots are required; removing heat from these sources limits the speed of imaging. For X-ray detectors, the intensity of the signal depends on the quantity of X-ray photons that reach the detector; all other things being equal, a detector with a smaller physical pixel size receives less photons than a detector with a larger physical pixel size, reducing the signal to noise ratio [71]. Both of these characteristics result in a micro-CT scanner with a slower scanning speed and lower signal to noise ratio than a comparable human CT scanner. The longer scan times do not allow for a complete acquisition of all projections around the animal in a single breath. If this is not considered during acquisition, each projection will be ac-

quired at a different stage of the respiratory cycle, resulting in artifacts in the CT image. To address this, respiratory gating, where each projection is acquired in the same part of the animal's respiratory cycle, is required.

5.1.1.1 Lack of contrast between lung parenchyma and soft tissue

The micro-CT scans have lower contrast between the lung parenchyma and soft tissue than in human CT scans. This is attributable to the amount of soft tissue in the lung parenchyma—the lung parenchyma is comprised primarily of the alveoli, which in a human, have a much larger mean linear intercept (MLI) of 210 μm compared to a mouse, 80 μm [72]. The MLI is a measure of the mean free distance between air spaces in the lung and is often used as an estimator of the alveolar diameter. The blood-gas barrier thickness in a human is 0.62 μm, compared to 0.32 μm for a mouse—these measurements suggest that a larger percentage of the lung parenchyma is tissue in a mouse lung compared to a human lung. Given the resolution of the micro-CT scan and the size of the alveoli, the lung parenchyma has a much higher radiodensity than the lung parenchyma in a human CT scan, which in turn reduces the contrast ratio between the lung parenchyma and other structures in the lungs. A comparison of histograms for the radiodensity of lung parenchyma and soft tissue for a human CT scan and murine micro-CT scan is shown in Figure 5.2. The histograms were generated by sampling ten manually selected 7x7x3 pixel regions in the lung parenchyma and six regions in various areas of soft tissue. The separate and narrow distributions of radiodensity for the lung parenchyma and soft tissue in the human CT scan allows for a fixed threshold to segment the two tissue types in the CT scan—due to the large separation, slight variances from scan to scan will not have a significant effect on the resulting segmentation. On the other hand, in the



(a) (b)

Figure 5.2: Comparison of the histograms of the radiodensity of lung parenchyma and soft tissue for a) a human CT scan and b) a murine micro-CT scan. Note the clear separation between the two tissue types in the human CT scan, while there is overlap in the murine scan.

micro-CT scan, the distributions of the lung parenchyma and soft tissue overlap. Not only does this overlap prevent a threshold from clearly segmenting the two tissue types, but the threshold has a significant effect on the segmentation of the two types.

5.1.1.2 Inconsistent scanner calibration

Care is taken to ensure that CT scanners used in clinical situations are properly calibrated; as a result, the intensity of a voxel of air can be expected to be -1000 HU and a voxel of water to be 0 HU. The micro-CT scanner used for the scans in this study was not well-calibrated; this is illustrated by the the histograms of the intensities of phantoms in Figure 5.4. There was a phantom, shown in Figure 5.3, with an equivalent density as air (-1000 HU), a water-equivalent phantom (0 HU), and a bone-equivalent phantom. The exact radiodensity of the bone

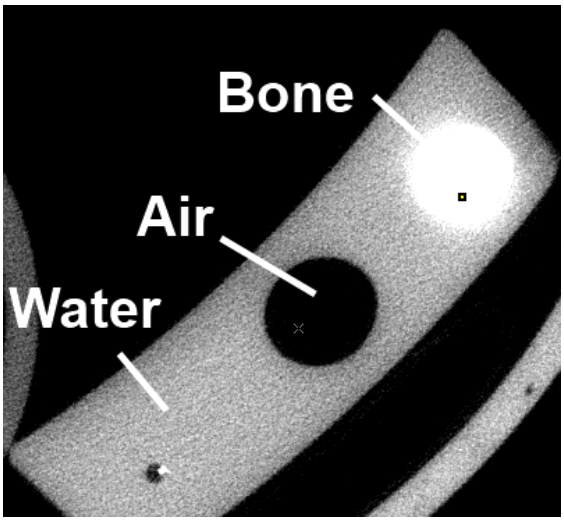
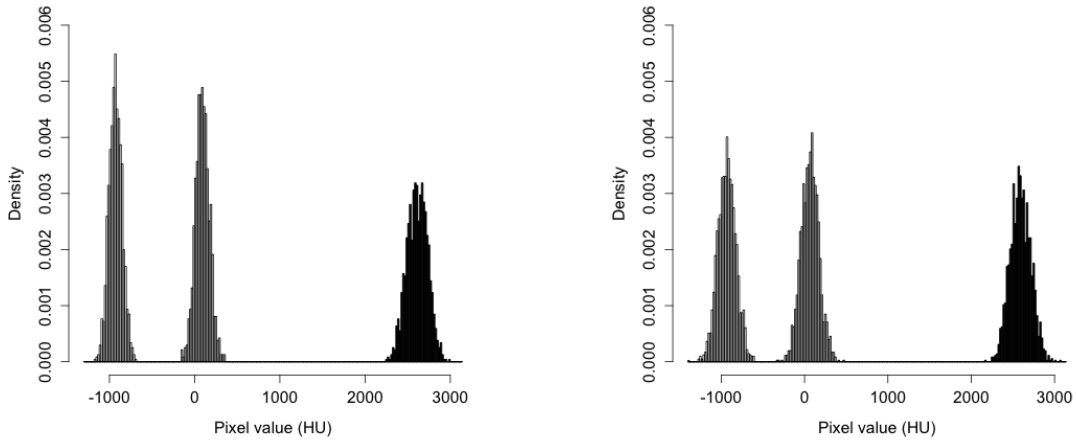


Figure 5.3: Phantom with material densities similar to water, air, and bone

phantom was unknown, but it should be the same in all scans. Note that the intensities in Figure 5.4a are shifted slightly lower than the intensities in Figure 5.4b, but the shift is not simply additive. The mean of the air, water, and bone intensity distributions are 4.2 HU, 25.7 HU, and 20.6 HU lower respectively in Figure 5.4b than in Figure 5.4a. In the ideal case, there should be no shift in the mean intensities of the phantoms between scans.

5.1.1.3 Live mouse imaging with micro-CT

Imaging a live mouse with a micro-CT scanner poses a challenge due to the lack of speed of the scanner and uncooperativeness of the mouse. These are both addressed by anesthetizing the mouse during the scan, which kept the mouse still and steadily breathing during the scan. The micro-CT scanner used in this study, a GE eXplore CT 120 micro-CT scanner, had limitations in the frequency of X-ray tube activations and the data processing speed. Configuring the scanner to produce a scan as fast as possible resulted in a scan time of multiple minutes for the field of view that included the entire mouse at the desired res-



(a) (b)

Figure 5.4: Histograms of intensities of phantoms of known radiodensity for two different scans. The materials, from left to right on the histograms, were air (-1000 HU), water (0 HU), and bone. In scan a), the mean intensities were -927.6 HU, 92.2 HU, and 2612.6 HU respectively, compared to scan b) with mean intensities of -931.8 HU, 66.5 HU, and 2592.0 HU respectively.

olution. Performing a scan in this configuration would have been unacceptable due to the respiratory motion of the mouse—the motion would cause blurring in the image.

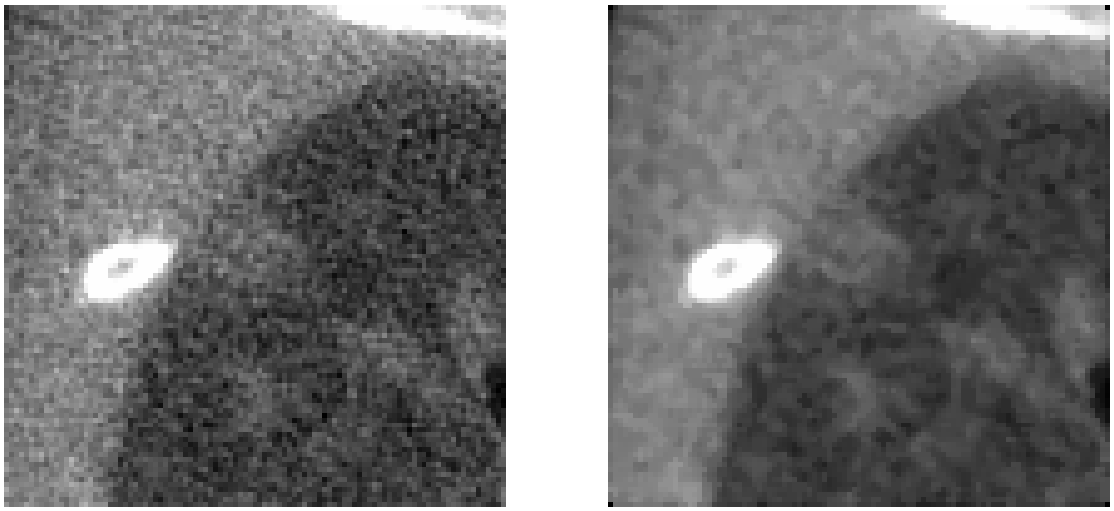
To address this problem, respiratory gating was used during the scans. In respiratory gating, the scanner is configured to only obtain image projections when the mouse is at the rest period in its respiratory cycle. Ideally, this would place the mouse lung in the same position and size for every projection. A view is a single position of the X-ray source around the gantry, and in this study, a scan had 720 projections, with 2 frames taken at each view to enable frame averaging.

5.1.2 Previous methods of pulmonary nodule measurement in mouse models

Most previous methods of pulmonary nodule measurement have been based on simple techniques, such as thresholding and region growing, with extensive manual intervention. Early approaches, such as Haines et al., did not attempt to specifically segment the pulmonary nodules; the soft tissue was segmented from the lung parenchyma using thresholds, region growing, and manual contouring, and the change in soft tissue volume was used as a measure of tumor burden [73]. Fushiki et al. used a similar method with additional manual and semi-automated segmentation to select only the tumors [74]. A more advanced method applied a semi-automated tool which presented a preliminary automated tumor boundary. However, manual modification of the boundary was required to produce an acceptable segmentation [75]. All of these methods require extensive user interaction with the system, which is time-consuming and may result in inconsistent measurements among users.

5.2 Pulmonary nodule growth rate algorithm modifications to support murine nodule measurement

The segmentation algorithm described in Chapter 1 was designed to segment pulmonary nodules from CT scans of human patients. As described in the previous section, there are many differences between a human CT scan and a murine micro-CT scan, requiring changes to the algorithm to accommodate these differences. The three main focus areas were the change in resolution, additional scanner noise, and reduced contrast in the micro-CT scans.



(a) (b)

Figure 5.5: The median filter, applied to the a) original image, reduces the visible noise in the b) output image

5.2.1 Pre-filtering and modified resampling to accommodate resolution change and noise

The noise in the micro-CT scan had a high enough intensity to interfere with the segmentation algorithm, primarily along the boundary between the nodule and lung parenchyma. This would manifest as either over- or under-segmentation of the nodule. To reduce this noise, a median filter was applied to the region of interest extracted from the micro-CT scan to remove the noise while maintaining the edges along the nodule. Based on empirical observations, a 2D median filter with a 3 x 3 pixel window was selected. Larger windows, as well as 3D windows, were tested, but did not have any visual improvement over the selected window size. The effect of the filter is illustrated in Figure 5.5, which shows a reduction in image noise in the output image.

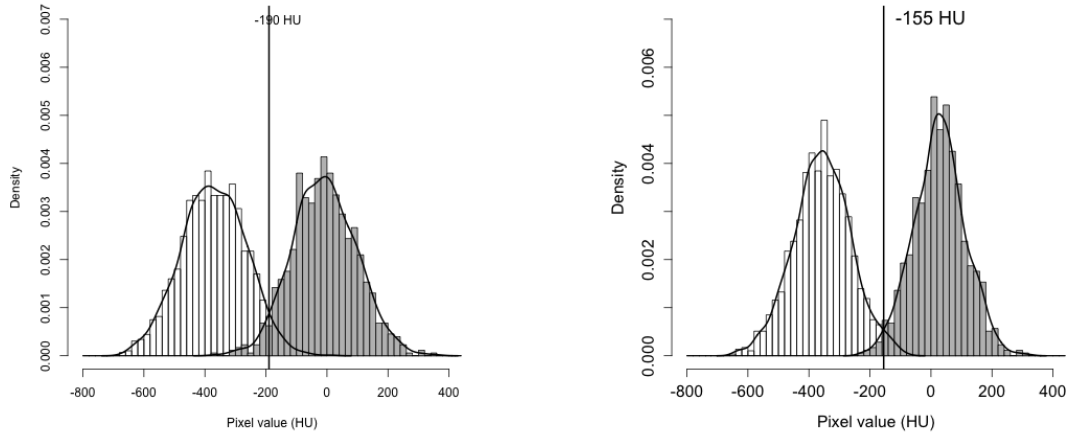
After the noise-reducing filter, the region of interested was resampled. In the original segmentation algorithm, the CT scan was resampled into 0.25 mm

isotropic space from a resolution of approximately 0.625 mm x 0.625 mm x 1.25 mm. For the micro-CT scans, which are imaged at an isotropic resolution of 50 μm, the images were resampled into an isotropic voxel space of 25 μm. The higher resolution provides for more precise boundaries along curves [22].

5.2.2 Adaptive thresholding for low contrast

The next step in the segmentation algorithm was to threshold the resampled region of interest to separate the soft tissue from the lung parenchyma. While the use of an adaptive threshold does not provide much benefit for segmenting nodules from human CT scans compared to a fixed threshold [23], murine micro-CT scans exhibit significant variation of the radiodensity of the lung parenchyma and soft tissue from one scan to another. The histograms of the radiodensity distributions of lung parenchyma and soft tissue are shown for two scans taken at two different times of the same mouse in Figure 5.6. Note that there is a bi-modal distribution, which indicates that we can separate the majority of the two tissue types with a threshold, but despite being the same mouse, these scans show a clear shift in the mean of both distributions. Using the same threshold for both scans would result in either over- or under-segmentation of the nodule. To compensate for this variation, an adaptive threshold was selected for each scan based on these histograms.

The histograms were generated by sampling ten manually selected 7x7x3 pixel regions in the lung parenchyma and six regions in various areas of soft tissue of the original micro-CT scans, prior to resampling or region of interest selection. The threshold was manually selected to be the midpoint of the two peaks, as shown by the plots in Figure 5.6 and provided as an input to the algo-



(a) (b)

Figure 5.6: Radiodensity distributions of the soft tissue and lung parenchyma in two different scans of the same mouse

rithm.

5.3 Exponential growth rate validation study

The goal of this study was to show that automated growth analysis was viable for micro-CT scans. However, verifying the correctness of the algorithm is difficult, due to many of the same factors described for human CT scans in Section 1.4. In brief, manual measurements are unreliable—the high inter- and intra-reader variability makes manual measurements unsuitable for ground truth. While we could have performed zero-change "coffee break"-type studies using mice, these studies have a significant limitation—measuring nodules with zero-change only provides information on how the algorithm performs in zero-change situations. To evaluate the algorithm for nodules with growth, we assumed the growth of nodules would follow the exponential growth model. If the nodules grow exponentially, the growth measured by the algorithm should

also show an exponential growth rate.

In this study, the growth rates of nodules in four mice were followed from the first appearance of the nodule in the scan until the mice began to show severe symptoms of disease. These growth rate measurements were computed by the automated algorithm described in Section 5.2 and by a manual measurement. Both sets of measurements were compared to the exponential growth model.

5.3.1 Murine dataset

This study used a novel mouse lung tumor model over-expressing the small subunit of the enzyme ribonucleotide reductase (RNR) combined with inactivation of the DNA repair gene Msh6; this model results in 100% of mice developing pulmonary nodules by six months of age [76]. These nodules histopathologically resemble human papillary adenocarcinomas, the most common form of human non-small cell lung cancer. All mice were maintained identically, following guidelines approved by the Cornell University Institutional Laboratory Animal Use and Care Committee. After the last micro-CT scans, the mice were euthanized by asphyxiation with carbon dioxide.

Nodules were selected that were present on at least three scans, were largely solid in appearance, and were only attached to the chest wall at a single point (the automated algorithm was only designed to handle a single attachment to a chest wall). Six nodules from four mice fulfilled these criteria and were followed in this study. Information about the mice is provided in Table 5.1. Each mouse was scanned a minimum of four times, with at least three weeks time interval in between. At least one tumor was visible for each mouse in all scans

	# of nodules	Age at initial scan (months)	Age at final scan (months)
1	1	11	14
2	1	12	15
3	2	12	17
4	2	11	15

Table 5.1: Information on mice in the study

at different time points. Nodules were identified on the each scan and followed in subsequent scans. The volume of each nodule at corresponding time point was computed by the automated algorithm and recorded for growth analysis. Two observers independently verified the segmentation of the nodules by the algorithm. At the end of the last live scan, mice were euthanized. Necropsy and histological analyses were performed to validate the nodules identified in micro-CT scans.

5.3.2 Scanning protocol

The protocol for scanning mice is complicated by the duration of a single scan, which was about 30–45 minutes, depending upon the breathing rate of the mouse. Prior to the scan acquisition process, the mice were anesthetized in an induction chamber with a continuous flow of 4% isoflurane/oxygen mixture. After the rate of breathing of the mice slowed, the isoflurane/oxygen mixture was reduced to 1-3% and maintained at this range during the duration of the micro-CT scan [77]. Scans of the chest region were acquired using a GE eXplore CT 120 micro-CT scanner with a tube current of 50 mA and a voltage of 100 kV. Each scan consisted of 720 projections in a single full rotation of the gantry. Two frames were acquired at each position of the gantry and averaged together prior to being transferred to the workstation for reconstruction; scans were reconstructed at 50 x 50 x 50 μm^3 voxel dimensions. Respiratory gating of the mice

was achieved using the BioVet physiological monitoring and triggering system (m2m Imaging Corp.). The final image volume varied according to the selected region of interest, but typically ranged from 400 x 400 x 500 slices to 800 x 550 x 1000 slices. Scans were converted from the manufacturer's proprietary format to DICOM and then imported into research software.

The scan parameters selected for this study were determined through empirical testing. Although previous studies have imaged mice using micro-CT, the scanner used in this study was newly released at the time of the study, so we needed to optimize the tradeoffs between speed, image quality, and radiation dosage of this particular scanner. The acceptable radiation exposure to the mouse and the scan time were the primary constraints that limited the quality of the scans. A single live mouse was scanned six times with different scanner parameters; the majority of these scans were performed with a low number of views per scan with no frame averaging, which substantially reduced the scan time to approximately 10 minutes. The time required for a scan directly increases with the number of views per scan and the number of frames used for averaging. The noise associated with each set of scanner parameters was quantified by measuring the standard deviations of the lung parenchyma radiodensity distributions, which were computed using the same approach described in Section 5.2.2, but with only five local regions. A table of the tested parameters and the resulting noise is provided in Table 5.2. The scan with the lowest noise was produced with the following parameters: 100 kVp, 50 mA, 20 ms exposure time, 360 projections, and 2 frame averaging. To increase the visible details, the parameters used in the following live imaging were 100 kVp, 50 mA, 20 ms exposure time, 720 projections and 2 frame averaging with 50μm reconstruction resolution. All live scans were acquired using these optimized parameters with

Table 5.2: Scan parameters and the associated noise measurements estimated using the standard deviation (SD) of the lung parenchyma radiodensity distribution

	Current (mA)	Voltage (kVp)	Exp. time (ms)	# proj.	Frame avg.	SD (HU)
1	32	70	32	220	1	53.3
2	50	70	32	220	1	80.4
3	32	80	32	220	1	45.6
4	50	80	32	220	1	41.1
5	50	100	32	220	1	39.2
6	50	100	20	360	2	30.5

Table 5.3: Micro-CT scanner parameters used in this study

	Current (mA)	Voltage (kVp)	Exp. time (ms)	# proj.	Frame avg.
All, except:	50	100	20	720	2
Mouse 1, T 3	50	100	20	720	1
Mouse 2, T 1	50	100	20	440	2

the exception of two scans shown in Table 5.3.

5.3.3 Growth rate evaluation

In this study, each mouse was scanned several times. The nodule growth rate can be computed from the nodule volume on each scan and time interval between scans. The exponential growth model is described in Section 1.3; some of the equations will be repeated here to provide context for the analysis.

The exponential growth model is defined for a pair of volume measurements as:

$$V_2 = V_1 \cdot e^{\lambda \cdot \Delta t} \tag{5.1}$$

where V_2 is the volume of the nodule at time t_2, V_1 is the volume of the nodule at time t_1, and Δt is the time interval $t_2 - t_1$ in days. Since we have more than two scans, we fit all of the measurements for a nodule to a single exponential model. This was performed using non-linear least squares regression in the R statistical package [68], which provides estimates of the parameters V_1 and λ. The goodness of fit of the model was quantified using the residual standard error (RSE); a perfect fit to the model would have an RSE of 0.

For a pair of scans, we typically use the growth index (GI) to report the growth rate:

$$\mathrm{GI} = 100 \cdot \left[\left(V_2 / V_1 \right)^{30.4375/\Delta t} - 1 \right] \tag{5.2}$$

We can also report a GI for a particular value of λ by rearranging Equation 5.1 in terms of the exponential coefficient λ:

$$\lambda = \frac{\ln\left(\frac{V_2}{V_1}\right)}{\Delta t} = \frac{\ln V_2 - \ln V_1}{t_2 - t_1} \tag{5.3}$$

which, after substitution into Equation 5.2, results in an expression in terms of λ:

$$\mathrm{GI}_{\mathrm{model}} = 100 \cdot \left[e^{30.4375\lambda} - 1 \right] \tag{5.4}$$

If the nodule growth follows the exponential model, as we hypothesize, the RSE will be low.

5.3.4 Comparison with manual measurements

The growth measured by the automated algorithm was compared to the growth manually measured by an observer—the author. The author measured each tumor by selecting the axial slice through the tumor with the largest cross-sectional area and marking the largest diameter. This diameter was converted

to a volume measurement, assuming a uniform spherical model; this closely mirrors the measurement method used in RECIST [9]. The growth rates from the manual measurements were evaluated as described in the previous section and compared to the automated measurements.

5.4 Exponential growth rate validation results

In this study, we monitored the progression of six pulmonary nodules presenting in four mice using micro-CT. The growth of each nodule over time, computed by the automated method, is illustrated in Figure 5.7. In this graph, the volume of the nodule is plotted over the time from the initial scan. The data were fit to an exponential growth model, and this model was used to estimate the growth index (GI) of the nodules, which are provided in Table 5.4. The initial nodule volume varied from 0.050 mm^3 to 0.898 mm^3. Nodules with smaller initial volumes tended to grow slightly faster, with the exception of the nodule in mouse 2. The GI values of the nodules ranged from 20.04–81.04% per month. Information on the model fits is provided in Table 5.6; the nodules with the lowest residual standard error (RSE) were 3B and 4A, if nodule 1, which only had three measurements, is discounted. These also appear to have the closest fit to the exponential model in the plot in Figure 5.7.

The manual growth measurements are plotted in Figure 5.8 and provided in Table 5.5; the GI values ranged from 24.12–205.58% per month. The initial volumes from the manual measurements were always larger than the automated measurements, with the exception of nodule 4B. This may have been due to the longest dimension of the nodule lying along the axial dimension of the scan. The growth rates were also larger for the manual measurements in three of the six

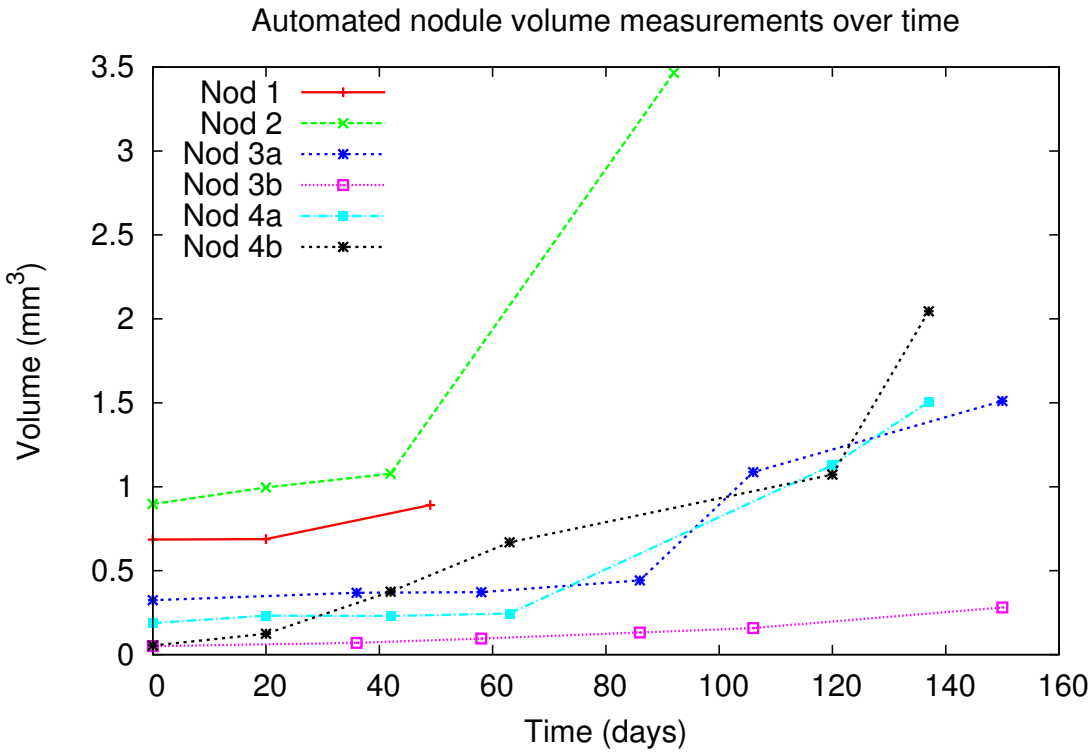


Figure 5.7: Plot of automated measurements over time for all nodules in the study

Table 5.4: Growth indices (GI) of six nodules detected in this study computed from automated volume measurements

Mouse (nodule)	Initial volume (mm^3)	Automated GI (%/month)
1	0.685	20.04
2	0.898	74.01
3 (A)	0.324	48.99
3 (B)	0.052	43.65
4 (A)	0.190	80.49
4 (B)	0.050	81.04

Table 5.5: Growth indices (GI) of six nodules detected in this study computed from manual volume measurements

Mouse (nodule)	Manual Initial Volume (mm^3)	Manual GI (%/month)
1	0.882	205.58
2	0.776	146.19
3 (A)	0.556	33.94
3 (B)	0.195	24.12
4 (A)	0.641	488.37
4 (B)	0.022	55.95

nodules, often by a significant amount. In the other three nodules, the growth rate was only slightly below the growth rate computed from the automated measurements. The fit of the data to the exponential model, provided in Table 5.7, was generally worse (higher RSE) for the manual measurements than the automated measurements with the exception of nodule 2, where the manual measurement of the volume on the last scan of nodule 2 showed a greater increase over the second to last scan than the automated measurement did. The growth rate computed from the manual measurements for nodule 4A had a large difference from the automated growth rate measurement and a much higher RSE.

5.5 Discussion

The hypotheses made for to this study were that the measured growth rate of murine pulmonary nodules was exponential and that the automated method would perform better than manual measurements, with measurements closer

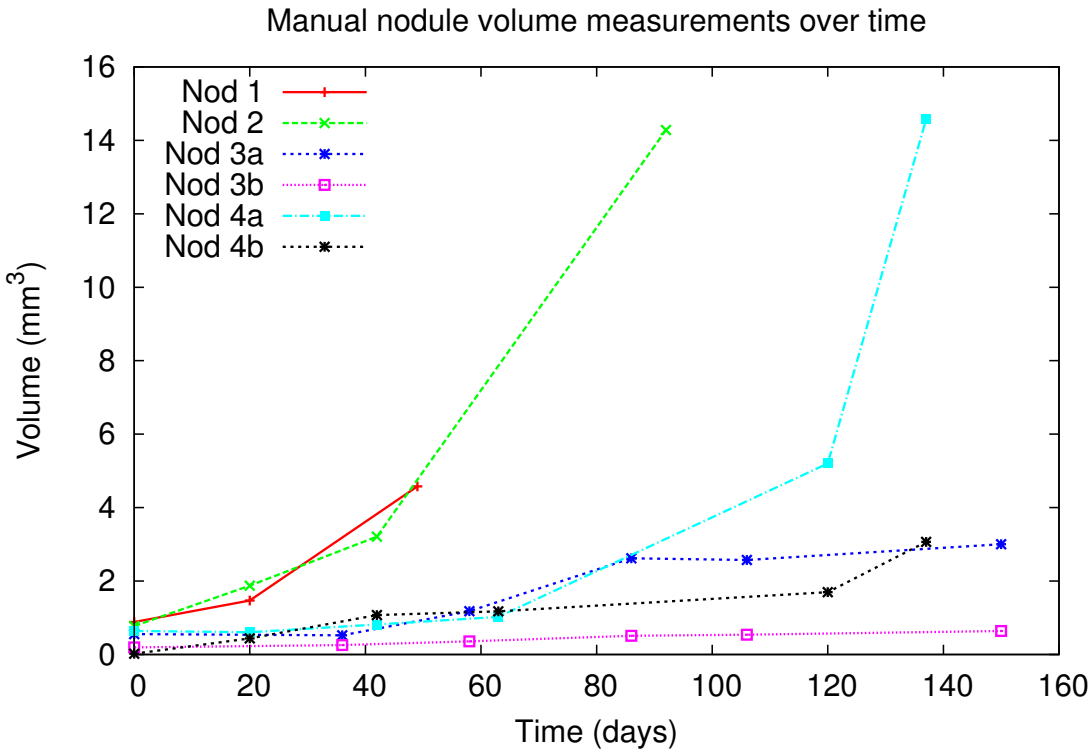


Figure 5.8: Plot of manual measurements over time for all nodules in the study

Table 5.6: Model fits for automated measurements (RSE = Residual standard error)

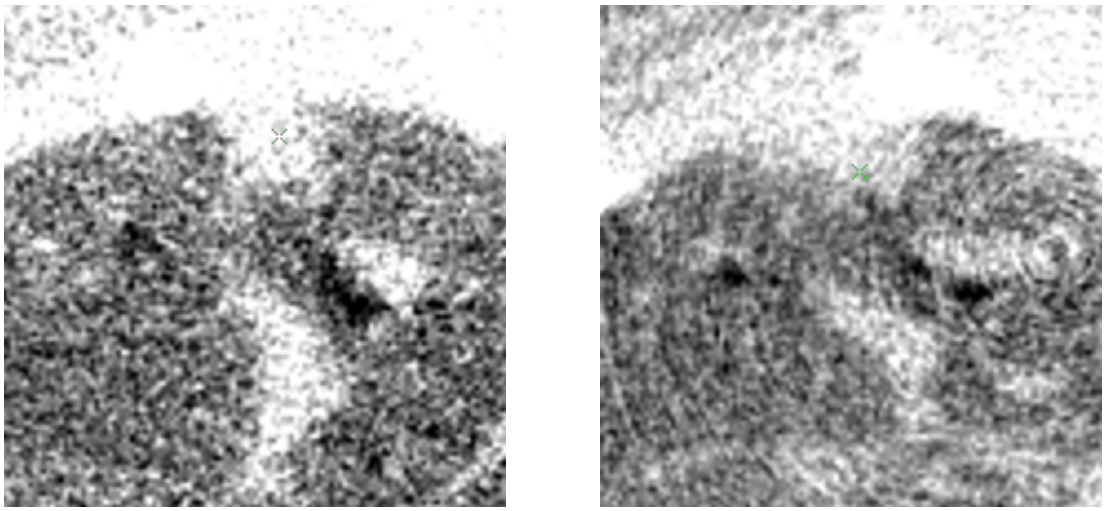
Mouse (Nodule)	# scans	Auto V_0	Auto λ	Auto RSE
1	3	0.6534	0.0060	0.0599
2	4	0.63854	0.0182	0.2850
3 (A)	6	0.2141	0.0131	0.1736
3 (B)	6	0.0466	0.0119	0.0044
4 (A)	6	0.1060	0.0194	0.0827
4 (B)	6	0.1310	0.0195	0.2067

Table 5.7: Model fits for manual measurements

Mouse (Nodule)	# scans	Manual V_0	Manual λ	Manual RSE
1	3	0.7541	0.0367	0.1659
2	4	0.9400	0.0296	0.1741
3 (A)	6	0.7918	0.0096	0.5851
3 (B)	6	0.2351	0.0071	0.0560
4 (A)	6	0.0050	0.0583	0.7166
4 (B)	6	0.3784	0.0146	0.3959

to the exponential model. The growth rates computed from both the automated and manual measurements show a trend towards exponential growth, as shown by the plots in Figures 5.7 and 5.8, and the growth measured by the automated method had better agreement with the exponential growth model, evidenced by the lower residual standard errors (RSE), than the manual method on five of the six nodules. Assuming the actual growth of the nodules followed the exponential model, this suggests that the automated measurements are more accurate.

The automated method measured smaller volumes than the manual method in all cases; this was due to the uni-dimensional manual measurement on the slice where the nodule appeared largest. The volume for the manual method was computed assuming a spherical model with a radius equal to half the uni-dimensional measurement, which would result in a sphere that totally encloses the nodule, unless the largest diameter of the nodule lies in the axial dimension. Nodules 1, 2, and 4A had higher GI values for the manual method, while the remaining nodules had higher GI values for the automated method. The large GI values for the manual method on these nodules are likely due to asymmetric growth of the nodule, which would be less likely to be captured correctly by a single uni-dimensional measurement. Regions of interest around nodule



(a) (b)

Figure 5.9: Region of interest of a single slice of the a) first scan and b) last scan for nodule 1 showing asymmetric growth along the chest wall

1 on the first and last scans are shown in Figure 5.9 showing the asymmetric growth of the nodule along the chest wall with very little growth perpendicular to the chest wall. The higher GI values for the automated method on the other three nodules was likely due to the increased sensitivity to change enabled by the use of all of the three-dimensional data instead of making a single linear measurement.

There is some deviation from the exponential growth model for all the nodules, which could be due to either an actual change in the nodule growth rate or sources of variation from the algorithm and data acquisition process. There is some evidence for non-exponential tumor growth, such as Gompertzian growth or a hybrid exponential and Gompertzian model [78], but in this case, we do not know the "true" growth of the nodule and so are unable to verify whether these deviations occur at the physiological level. Both the manual and automated measurements show deviation from the exponential growth model, which suggests that these deviations are not specific to either method.

The process of acquiring the micro-CT scans from the live mice had several sources of variation that may have affected the measurements from those scans. As discussed in Section 5.1.1, not only were the micro-CT scans were not well-calibrated from one scan to another, but there was overlap between the intensity histograms of the lung parenchyma and soft tissue. These factors made the placement of the boundary between the nodule and lung parenchyma difficult for both the human observer and the algorithm. In the automated method, we used an adaptive threshold that was calculated in the same manner in every scan, but any variations in this threshold would have resulted in an apparent change in size beyond the actual nodule size change. Some of the variation could be removed in future scans by performing more frequent scanner calibrations and including phantoms with known densities which could be used to compensate for any drift in the scanner calibration.

The protocol required for scanning live mice may have also caused some of the observed deviation from the exponential model. The shape of the lung was different in each scan due to the respiratory gating equipment used in this study. The respiratory gating was accomplished using a pressure sensor placed under the mouse; when the mouse took a breath, pressure would be applied to the sensor. This required external pressure to be applied to the mouse to ensure it was snug against the sensor—this pressure altered the shape of the chest cavity and may have compressed the lung. The mice also had more difficulty breathing in later scans, due to the progression of disease. To measure this change in the lung volume from scan to scan, the threshold found during the adaptive thresholding step of the algorithm was used to segment the lung area. Morphological filtering and connected component analysis was performed to isolate the lungs. We measured the lung volume in each scan for all the mice and found

that the change from the maximum lung volume to the minimum lung volume for a mouse ranged from 28% to 61%. This suggests that changes in the lung volume and morphology due to changes in mouse positions and disease burden may have contributed to variations in tumor volume measurements. This source of variation could be eliminated in future studies by using a ventilator to force the mouse to breathe at a fixed rate, ensuring that the lung is inflated to the same volume each time and eliminating the external pressure on the mouse for the pressure sensor.

5.6 Summary

In this study, we developed an automated pulmonary nodule segmentation algorithm for measuring murine tumors imaged by micro-CT. This algorithm was used to monitor the progression of six nodules in four mice. The growth of the nodules show agreement with the exponential growth model; however, there were deviations from a perfect exponential growth model, which may be due to variations in the scan calibration and changes in the mouse lung volumes. Nonetheless, the automated algorithm is able to carry out accurate measurements of nodule volumes and can be used to monitor disease progression, enabling the use of nodule volume measurements from micro-CT as an imaging biomarker in preclinical studies. In the future, this automated algorithm will be evaluated with a larger cohort of animals and improved to handle nodules with more difficult morphology. It can also be used to monitor disease progression upon drug treatment and test potential therapeutic responses.

CHAPTER 6

CONCLUSIONS

The primary goal of this dissertation was to improve the measurement of pulmonary nodule growth rates CT scans to provide radiologists with better information for making decisions regarding the follow-up and treatment of patients. The usefulness of growth rate measurements is limited by their uncertainty due to errors and measurement variation—reducing these would allow radiologists to quantify nodule growth earlier and improve patient outcomes.

A multifaceted approach was undertaken to advance the field of pulmonary nodule growth rate measurement. The performance of current algorithms was benchmarked against a common dataset for the first time. Two methods were developed to reduce growth measurement variation, and a semi-automated method was used to measure murine pulmonary nodule growth in micro-CT scans. This work had the following novel contributions:

- A standard dataset of pulmonary nodules was made publicly available and used to compare 18 different algorithms in the VOLCANO study [55].
- A moment-based method, ZCOMP, developed specifically to reduce the impact of additional uncertainty in the z-direction of modern CT scanners [62].
- A density-based nodule growth rate measurement method that improved diagnostic accuracy, especially for nodules with complex shapes [61].
- A semi-automated method to measure murine pulmonary nodule growth in micro-CT scans that was validated through comparison with the exponential growth model, allowing for growth rate measurement in preclinical studies of small animals.

6.1 VOLCANO study of algorithms on a standard dataset

Prior to the VOLCANO study [55], pulmonary nodule measurement algorithms were evaluated on different datasets using different metrics, making comparisons between methods difficult. Furthermore, most evaluations were performed using nodule size instead of nodule growth rate, which is the actual indicator used for diagnosis. A dataset was created for the VOLCANO study with 49 nodules, including both zero-change and growing nodules, and a single phantom, that was made publicly available. The growth rate results from eighteen different methods were analyzed to identify trends amongst the methods.

The variability between methods was similar for those nodules imaged using the same slice thickness, but differed for nodules on different thickness scans, implying that maintaining the same slice thickness is essential for reducing measurement variability. The variation between the methods for the zero-change nodules and those with change was similar, allowing zero-change nodules to be used for measuring the variation of methods. Finally, although the growth rate measurements were largely in agreement, the volume measurements provided by twelve methods were not—these differing results indicate that the results on volume can not be extrapolated to growth and vice versa.

6.2 ZCOMP method to address asymmetric growth in the z-direction

During a review of zero-change nodules, some nodules were observed on a different number of CT scan slices between scans that were only minutes apart. This additional variation in the z (axial) direction caused a bias in the growth rate measurement which was not addressed by any previous methods. The

ZCOMP algorithm [62] was developed to reduce the impact of any linear distortion in growth in the z-direction. The 3D image moments were calculated from a binary segmented image and used to establish a surrogate volume measurement. This ZCOMP method was evaluated on a dataset of 22 zero-change nodules and compared to a semi-automated volumetric segmentation method. Using ZCOMP, the variation was reduced from 95% limits of agreement of (-52.1%, 30.1%) to (-34.2%, 23.3%) for nodules imaged on both same-slice thickness and different-slice thickness scans.

6.3 Density-based growth rate measurement

The majority of nodule measurement algorithms segment the nodule from the lung parenchyma and other attached structures; however, some nodules have complex shapes and attachments that make explicit segmentation impossible. These nodules also tend to be malignant, so measuring their growth rate is important. To measure these nodules, the growth index from density (GI_D) method was developed [61]. The underlying idea of the GI_D method was that, as a nodule grows, more cells are added, which should increase the density of the nodule; thus, the change in density in a region of interest around the nodule was used as a surrogate growth measure. In order to ensure the regions from one scan to the next were as similar as possible, a rigid registration was performed to align the two regions of interest.

On a dataset of 20 zero-change nodules, the GI_D method exhibited nearly half the variation, (-19.7%,15.3%), of a volumetric segmentation method, (-54.6%, 30.3%). The diagnostic performance, measured on a dataset of 38 stable and 23 malignant nodules (4 of which were nodules with complex shapes), improved from 74% correct for malignant nodules for the volumetric method to 96% for

the GI_D method. Much of this improvement came from complex nodules that were incorrectly segmented by the volumetric method.

6.4 Semi-automated growth measurement of murine pulmonary nodules

While there have been many methods for measuring pulmonary nodule growth for human pulmonary nodules, the methods available for murine pulmonary nodule growth measurement are much cruder, requiring extensive manual intervention. Murine models are extensively used in pre-clinical research, where quantifying disease progression and treatment response are important. The quality of micro-CT scans used in these pre-clinical study poses three main challenges: additional noise, lack of contrast between the lung parenchyma and soft tissue, and poor calibration. A semi-automated volumetric method was adapted to address these issues. One of the most significant changes was the use of adaptive thresholding. In contrast to human CT scans, adaptive thresholding was necessary to ensure accurate segmentation on micro-CT scans, due to the shift in mean intensity cause by the poor calibration of the scanner, as well as an overlap between the density histograms of the two tissue types. To evaluate the method, instead of relying on manual measurements as truth, the nodule growth rate was compared to the exponential growth model. Manual measurements were also included and compared to the exponential growth model.

Six nodules from four mice were identified for this study. The semi-automated method measured growth that was a better fit to the exponential model than the manual method for five of the six nodules and only slightly worse for the remaining nodule. This suggests that the semi-automated method was able to

measure the murine pulmonary nodules better than manual measurements in this study.

6.5 Future work

There are several logical extensions to the work presented here. The dataset created for VOLCANO was successful in evaluating automated methods, but including manual measurements by radiologists would allow for direct comparisons between automated and manual methods. Radiologist performance and variation could be characterized in the different subgroups of nodules to better understand the behavior of manual measurements in different situations. Also, expanding the dataset to include nodules with known diagnoses would allow for the evaluation of classification performance from growth rate measurements.

There are many opportunities in pre-clinical studies for the development and application of automated nodule measurement methods. Quantification of the variation in volume and growth measurement is just as important an issue in pre-clinical studies as with human patients—better measurements would enable earlier identification of disease or treatment response. The variation of both manual and automated methods could be characterized. The preliminary work on measuring murine pulmonary nodules could be extended to nodules in other organs, such as the liver, in order to provide a reliable, non-manual method of quantifying disease and treatment response for other forms of cancer.

BIBLIOGRAPHY

[1] V. P. Collins, R. K. Loeffler, and H. Tivey, "Observations on growth rates of human tumors," *American Journal of Roentgenology*, vol. 76, pp. 988–1000, November 1956.

[2] M. H. Nathan, V. P. Collins, and R. A. Adams, "Differentiation of benign and malignant pulmonary nodules by growth rate," *Radiology*, vol. 79, pp. 221–231, August 1962.

[3] American Cancer Society, *Cancer Facts & Figures 2012*. Atlanta: American Cancer Society, 2012.

[4] C. I. Henschke, D. F. Yankelevitz, D. M. Libby, M. W. Pasmantier, J. P. Smith, and O. S. Miettinen, "Survival of patients with stage I lung cancer detected on CT screening.," *The New England journal of medicine*, vol. 355, pp. 1763–71, Oct. 2006.

[5] N. C. Institute, "Lung cancer trial results show mortality benefit with low-dose CT." Press release, Nov. 2010.

[6] D. M. Libby, J. P. Smith, N. K. Altorki, M. W. Pasmantier, D. Yankelevitz, and C. I. Henschke, "Managing the Small Pulmonary Nodule Discovered by CT," *Chest*, vol. 125, no. 4, pp. 1522–1529, 2004.

[7] H. MacMahon, J. H. M. Austin, G. Gamsu, C. J. Herold, J. R. Jett, D. P. Naidich, E. F. Patz, and S. J. Swensen, "Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society.," *Radiology*, vol. 237, pp. 395–400, Nov. 2005.

[8] C. R. Meyer, C. P. Fenimore, G. Mclennan, L. M. Bidaut, D. P. Barboriak, M. A. Gavrielides, E. F. Jackson, M. F. Mcnitt-gray, P. E. Kinahan, N. Petrick, and B. Zhao, "Quantitative Imaging to Assess Tumor Response to Therapy: Common Themes of Measurement, Truth Data, and Error Sources," *Translational Oncology*, vol. 2, no. 4, pp. 198–210, 2009.

[9] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *European journal of cancer*, vol. 45, pp. 228–47, Jan. 2009.

[10] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Bean CT*, ch. 2. Berlin: Springer, 2008.

[11] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Bean CT*, ch. 3. Berlin: Springer, 2008.

[12] J. Rydberg, K. A. Buckwalter, K. S. Caldemeyer, M. D. Phillips, D. J. Conces, A. M. Aisen, S. A. Persohn, and K. K. Kopecky, "Multisection CT: scanning techniques and clinical applications.," *Radiographics*, vol. 20, no. 6, pp. 1787–806, 2000.

[13] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Bean CT*, ch. 5. Berlin: Springer, 2008.

[14] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Bean CT*, ch. 8. Berlin: Springer, 2008.

[15] R. A. Brooks and G. D. Chiro, "Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging," *Physics in Medicine and Biology*, vol. 21, no. 5, pp. 689–732, 1976.

[16] D. J. Brenner and E. J. Hall, "Computed tomography–an increasing source of radiation exposure," *The New England Journal of Medicine*, vol. 357, pp. 2277–84, Nov. 2007.

[17] J. G. Ravenel, W. M. Leue, P. J. Nietert, J. V. Miller, K. K. Taylor, and G. a. Silvestri, "Pulmonary nodule volume: effects of reconstruction parameters on automated measurements–a phantom study.," *Radiology*, vol. 247, pp. 400–8, May 2008.

[18] C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *AJR. American journal of roentgenology*, vol. 178, pp. 1053–7, May 2002.

[19] J. F. Barrett and N. Keat, "Artifacts in CT: recognition and avoidance," *Radiographics*, vol. 24, no. 6, pp. 1679–91, 2004.

[20] W. J. Kostis, D. F. Yankelevitz, A. P. Reeves, S. C. Fluture, and C. I. Henschke, "Small pulmonary nodules: Reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT," *Radiology*, vol. 231, pp. 446–452, May 2004.

[21] G. A. Lillington, "Management of solitary pulmonary nodules," *Disease-a-Month*, vol. 37, pp. 269–318, May 1991.

[22] W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 1259–1274, October 2003.

[23] A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 435–450, April 2006.

[24] J. P. Ko, H. Rusinek, E. L. Jacobs, J. S. Babb, M. Betke, G. McGuinness, and D. P. Naidich, "Small pulmonary nodules: Volume measurement at chest CT – phantom study," *Radiology*, vol. 228, pp. 864–870, September 2003.

[25] J.-M. Kuhnigk, V. Dicken, L. Bornemann, A. Bakai, D. Wormanns, S. Krass, and H.-O. Peitgen, "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans.," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 417–34, Apr. 2006.

[26] R. Opfer and R. Wiemker, "A new general tumor segmentation framework based on radial basis function energy minimization with a validation study on LIDC lung nodules," in *Proceedings of SPIE*, vol. 6512, pp. 651217–651217–10, SPIE, 2007.

[27] J. Wang, R. Engelmann, and Q. Li, "Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique," *Medical Physics*, vol. 34, no. 12, pp. 4678–4689, 2007.

[28] K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT," *IEEE Transactions on Medical Imaging*, vol. 24, pp. 409–423, March 2005.

[29] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, K. Yamada, H. Nishiyama, K. Eguchi, M. Kaneko, and N. Moriyama, "A computerized approach for estimating pulmonary nodule growth rates in three-dimensional thoracic CT images based on CT density histogram," in *Medical Imaging 2005: Image Processing* (J. M. Fitzpatrick and J. M. Reinhardt, eds.), vol. 5747, pp. 872–882, SPIE, 2005.

[30] S. G. Jennings, H. T. Winer-Muram, R. D. Tarver, and M. O. Farber, "Lung

tumor growth: Assessment with CT – comparison of diameter and cross-sectional area with volume measurements," *Radiology*, vol. 231, pp. 866–871, June 2004.

[31] D. Wormanns, S. Diederich, M. Lentschig, F. Winter, and W. Heindel, "Spiral CT of pulmonary nodules: interobserver variation in assessment of lesion size," *Eur. Radiol.*, vol. 10, pp. 710–713, 2000.

[32] N. R. Bogot, E. a. Kazerooni, A. M. Kelly, L. E. Quint, B. Desjardins, and B. Nan, "Interobserver and intraobserver variability in the assessment of pulmonary nodule size on CT using film and computer display methods," *Academic Radiology*, vol. 12, pp. 948–56, Aug. 2005.

[33] M.-P. Revel, A. Bissery, M. Bienvenu, L. Aycard, C. Lefort, and G. Frija, "Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable?," *Radiology*, vol. 231, pp. 453–8, May 2004.

[34] J. J. Erasmus, G. W. Gladish, B. S. Sabloff, L. Broemeling, M. T. Truong, R. S. Herbst, and R. F. Munden, "Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 21, pp. 2574–82, July 2003.

[35] A. M. Biancardi, A. P. Reeves, D. F. Yankelevitz, D. Ghiorghiu, M. Scott, and H. Mann, "A pilot study evaluating pulmonary nodule marking methods," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 85–88, IEEE, June 2009.

[36] J. C. Ross, J. V. Miller, W. D. Turner, and T. P. Kelliher, "An analysis of early studies released by the Lung Imaging Database Consortium (LIDC)," *Academic radiology*, vol. 14, no. 11, pp. 1382–8, 2007.

[37] A. P. Reeves, A. M. Biancardi, T. V. Apanasovich, C. R. Meyer, H. MacMahon, E. J. van Beek, E. A. Kazerooni, D. Yankelevitz, M. F. McNitt-Gray, G. McLennan, S. G. Armato III, C. I. Henschke, D. R. Aberle, B. Y. Croft, and L. P. Clarke, "The lung image database consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements," *Academic Radiology*, vol. 14, pp. 1475–1485, Dec 2007.

[38] M. a. Gavrielides, L. M. Kinnard, K. J. Myers, R. Zeng, and N. Petrick, "FDA phantom CT database: a resource for the assessment of lung nodule size estimation methodologies and software development," in *Medical Imaging*

2010: Computer-Aided Diagnosis (N. Karssemeijer and R. M. Summers, eds.), vol. 7624, (San Diego CA USA), pp. 762417–762417–8, SPIE, 2010.

[39] D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: Volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, pp. 251–256, October 2000.

[40] M. Das, G. Muhlenbruch, M. Katoh, A. Bakai, M. Salganicoff, S. Stanzel, A. H. Mahnken, R. W. Gunther, and J. E. Joachim, "Automated volumetry of solid pulmonary nodules in a phantom: Accuracy across different CT scanner technologies," *Investigative Radiology*, vol. 45, pp. 297–302, May 2007.

[41] M. Das, J. Ley-Zaporozhan, H. a. Gietema, A. Czech, G. Mühlenbruch, A. H. Mahnken, M. Katoh, A. Bakai, M. Salganicoff, S. Diederich, M. Prokop, H.-U. Kauczor, R. W. Günther, and J. E. Wildberger, "Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners," *European Radiology*, vol. 17, pp. 1979–84, Aug. 2007.

[42] A. Marchianò, E. Calabrò, E. Civelli, G. Di Tolla, L. F. Frigerio, C. Morosi, F. Tafaro, E. Ferri, N. Sverzellati, T. Camerini, L. Mariani, S. Lo Vullo, and U. Pastorino, "Pulmonary nodules: volume repeatability at multidetector CT lung cancer screening," *Radiology*, vol. 251, pp. 919–25, June 2009.

[43] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement.," *Lancet*, vol. 1, pp. 307–10, Feb. 1986.

[44] D. Wormanns, G. Kohl, E. Klotz, A. Marheine, F. Beyer, W. Heindel, and S. Diederich, "Volumetric measurements of pulmonary nodules at multirow detector CT: In vivo reproducibility," *European Radiology*, vol. 14, pp. 86–92, 2004.

[45] L. R. Goodman, M. Gulsun, L. Washington, P. G. Nagy, and K. L. Piacsek, "Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements," *American Journal of Roentgenology*, vol. 186, pp. 989–994, April 2006.

[46] H. A. Gietema, C. M. Schaefer-Prokop, W. P. T. M. Mali, G. Groenewegen, and M. Prokop, "Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT influence of inspiration level, nodule size, and segmentation performance," *Radiology*, vol. 245, no. 3, pp. 888–894, 2007.

[47] B. Zhao, L. P. James, C. S. Moskowitz, P. Guo, M. S. Ginsberg, R. A. Lefkowitz, Y. Qin, G. J. Riely, M. G. Kris, and L. H. Schwartz, "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer," *Radiology*, vol. 252, pp. 263–72, July 2009.

[48] C. Rampinelli, E. De Fiori, S. Raimondi, G. Veronesi, and M. Bellomi, "In vivo repeatability of automated volume calculations of small pulmonary nodules with CT," *American Journal of Roentgenology*, vol. 192, no. 6, pp. 1657–61, 2009.

[49] Y. Wang, R. J. van Klaveren, H. J. van der Zaag-Loonen, G. H. de Bock, H. A. Gietema, D. M. Xu, A. L. M. Leusveld, H. J. de Koning, E. T. Scholten, J. Verschakelen, M. Prokop, and M. Oudkerk, "Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program," *Radiology*, vol. 248, pp. 625–31, Aug. 2008.

[50] M. Petrou, L. E. Quint, B. Nan, and L. H. Baker, "Pulmonary nodule volumetric measurement variability as a function of ct slice thickness and nodule morphology," *Am. J. Roentgenol.*, vol. 188, pp. 306–312, Feb. 2007.

[51] J. P. Ko, R. Marcus, E. Bomsztyk, J. S. Babb, C. Stefanescu, M. Kaur, D. P. Naidich, and H. Rusinek, "Effect of blood vessels on measurement of nodule volume in a chest phantom," *Radiology*, vol. 239, pp. 79–85, April 2006.

[52] B. Zhao, L. Schwartz, C. Moskowitz, L. Wang, M. Ginsberg, C. Cooper, L. Jiang, and J. Kalaigian, "Pulmonary metastases: Effect of CT section thickness on measurement – initial experience," *Radiology*, vol. 234, pp. 934–939, March 2005.

[53] J. G. Ravenel, W. M. Leue, P. J. Nietert, J. V. Miller, K. K. Taylor, and G. a. Silvestri, "Pulmonary nodule volume: effects of reconstruction parameters on automated measurements–a phantom study," *Radiology*, vol. 247, pp. 400–8, May 2008.

[54] A. C. Jirapatnakul, Y. D. Mulman, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Segmentation of juxtapleural pulmonary nodules using a robust surface estimate.," *International journal of biomedical imaging*, vol. 2011, p. 632195, Jan. 2011.

[55] A. P. Reeves, A. C. Jirapatnakul, A. M. Biancardi, T. V. Apanasovich, C. Schaefer, J. J. Bowden, M. Kietzmann, R. Korn, M. Dillmann, Q. Li,

J. Wang, J. H. Moltz, J.-M. Kuhnigk, T. Hayashi, X. Zhou, H. Fujita, T. Duindam, R. Avila, J. P. Ko, K. Melamud, H. Rusinek, R. Wiemker, G. Soza, C. Tietjen, M. Thorn, M. F. McNitt-Gray, Y. Valenciaga, M. Khatonabadi, Y. Kawata, and N. Niki, "The VOLCANO'09 Challenge: Preliminary Results," in *Second International Workshop on Pulmonary Image Analysis* (M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.-M. Kuhnigk, C. Lorenz, J. McClelland, K. Mori, A. Reeves, and J. Reinhardt, eds.), (London), pp. 353–364, MICCAI, 2009.

[56] A. P. Reeves, A. M. Biancardi, D. Yankelevitz, S. Fotin, B. M. Keller, A. Jirapatnakul, and J. Lee, "A public image database to support research in computer aided diagnosis.," in *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2009, pp. 3715–8, Jan. 2009.

[57] T. Duindam, B. D. Hoop, and B. V. Ginneken, "Nodule volume change estimation in thoracic CT using sphere fitting , morphological segmentation and image registration," in *The Second International Workshop on Pulmonary Image Analysis*, (London, UK), pp. 389–400, 2009.

[58] I. C. Sluimer, M. Prokop, I. Hartmann, and B. van Ginneken, "Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution ct of the lung," *Medical Physics*, vol. 33, no. 7, pp. 2610–2620, 2006.

[59] T. Hayashi, X. Zhou, and H. Fujita, "Automated approach to measure pulmonary nodule volume based on radius and CT number," in *The Second International Workshop on Pulmonary Image Analysis*, no. 1, (London, UK), pp. 377–388, 2009.

[60] M. Athelogou, G. Schmidt, A. Schäpe, M. Baatz, and G. Binnig, *Cognition Network Technology–A Novel Multimodal Image Analysis Technique for Automatic Identification and Quantification of Biological Image Contents*, pp. 407–422. Springer, 2007.

[61] A. C. Jirapatnakul, A. P. Reeves, A. M. Biancardi, D. F. Yankelevitz, and C. I. Henschke, "Semi-automated measurement of pulmonary nodule growth without explicit segmentation," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI 2009. IEEE International Symposium on*, pp. 855–858, June 2009.

[62] A. C. Jirapatnakul, A. P. Reeves, A. M. Biancardi, D. F. Yankelevitz, and C. I. Henschke, "Improved precision of repeat image change measurement

of pulmonary nodules using moment-based z-compensation on a zero-change dataset," *Proceedings of SPIE*, pp. 72601P–72601P–9, 2009.

[63] A. P. Reeves, R. J. Prokop, S. E. Andrews, and F. P. Kuhl, "Three dimensional shape analysis using moments and fourier descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 937–943, 1988.

[64] A. C. Jirapatnakul, A. P. Reeves, T. V. Apanasovich, M. D. Cham, D. F. Yankelevitz, and C. I. Henschke, "Characterization of solid pulmonary nodules using three-dimensional features," *SPIE International Symposium on Medical Imaging 2007*, Feb. 2007.

[65] A. C. Jirapatnakul, A. P. Reeves, T. V. Apanasovich, A. M. Biancardi, D. F. Yankelevitz, and C. I. Henschke, "Characterization of pulmonary nodules: Effects of size and feature type on reported performance," in *SPIE International Symposium on Medical Imaging*, vol. 6915, p. 69151E, Feb 2008.

[66] R. D. Millán, L. Dempere-Marco, J. M. Pozo, J. R. Cebral, and A. F. Frangi, "Morphological characterization of intracranial aneurysms using 3-d moment invariants," *IEEE Transactions on Medical Imaging*, vol. 26, no. 9, pp. 1270–1282, 2007.

[67] J. W. Eaton, *GNU Octave Manual*. Network Theory Limited, 2002.

[68] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[69] K. K. Frese and D. a. Tuveson, "Maximizing mouse cancer models.," *Nature Reviews Cancer*, vol. 7, pp. 645–58, Sept. 2007.

[70] M. J. Paulus, S. S. Gleason, S. J. Kennel, P. R. Hunsicker, and D. K. Johnson, "High resolution X-ray computed tomography: an emerging tool for small animal cancer research.," *Neoplasia (New York, N.Y.)*, vol. 2, no. 1-2, pp. 62–70, 2008.

[71] M. J. Yaffe and J. a. Rowlands, "X-ray detectors for digital radiography.," *Physics in medicine and biology*, vol. 42, pp. 1–39, Jan. 1997.

[72] C. G. Irvin and J. H. T. Bates, "Measuring the lung function in the mouse: the challenge of size.," *Respiratory research*, vol. 4, p. 4, Jan. 2003.

[73] B. B. Haines, K. A. Bettano, M. Chenard, R. S. Sevilla, C. Ware, M. H. Angagaw, C. T. Winkelmann, C. Tong, J. F. Reilly, C. Sur, and W. Zhang, "A quantitative volumetric micro-computed tomography method to analyze lung tumors in genetically engineered mouse models.," *Neoplasia*, vol. 11, pp. 39–47, Jan. 2009.

[74] H. Fushiki, T. Kanoh-Azuma, M. Katoh, K. Kawabata, J. Jiang, N. Tsuchiya, A. Satow, Y. Tamai, and Y. Hayakawa, "Quantification of mouse pulmonary cancer models by microcomputed tomography imaging.," *Cancer science*, pp. 1–6, May 2009.

[75] D. D. Cody, C. L. Nelson, W. M. Bradley, M. Wislez, D. Juroske, R. E. Price, X. Zhou, B. N. Bekele, and J. M. Kurie, "Murine lung tumor measurement using respiratory-gated micro-computed tomography.," *Investigative radiology*, vol. 40, pp. 263–9, May 2005.

[76] X. Xu, J. L. Page, J. A. Surtees, H. Liu, S. Lagedrost, Y. Lu, R. Bronson, E. Alani, A. Y. Nikitin, and R. S. Weiss, "Broad overexpression of ribonucleotide reductase genes in mice specifically induces lung neoplasms.," *Cancer research*, vol. 68, pp. 2652–60, Apr. 2008.

[77] C.-S. Hung, K. W. Dodson, and S. J. Hultgren, "A murine model of urinary tract infection.," *Nature protocols*, vol. 4, pp. 1230–43, Jan. 2009.

[78] M. W. Retsky, D. E. Swartzendruber, R. H. Wardwell, and P. D. Bame, "Is Gompertzian or exponential kinetics a valid description of individual human cancer growth?," *Medical hypotheses*, vol. 33, pp. 95–106, Oct. 1990.