

An Automated System to Assess Air Space Size in Histopathology Images of Lung Tissue

by

Diego Politis

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Masters of Applied Science

in

Biomedical Engineering

Carleton University
Ottawa, Ontario

© 2022, Diego Politis

Abstract

The mean linear intercept (MLI) score is a useful and common approach for quantifying lung structure in histopathological images. This thesis describes a system developed to calculate the MLI score in a fully automated manner. The system was tested using 20 WSIs from mice. The root-mean-squared deviation between the MLI score of the proposed method and a human rater was 5.73 (standard deviation 5.65), and there was a very strong correlation ($r=0.9931$). Biases for the indirect method of MLI scoring are examined and shown to account for the differences with the direct MLI scores. Results suggest that shorter guideline length and smaller number of accepted FOV images have a higher standard error for MLI scoring when compared to longer guideline lengths and higher number of accepted FOV images. The proposed automated system provides an efficient, accurate, and accessible method that could replace current manual and semi-automated techniques.

Acknowledgements

An incredible journey has come to an end, one that I feel so fortunate to have embarked on. This accomplishment is the culmination of efforts from numerous people who have shaped me into the person I am today. I feel fortunate to be able to express my gratitude to the individuals who helped make this achievement a reality.

I first want to express my sincere appreciation to my supervisor, Dr. Adrian Chan, for his guidance and mentorship throughout my thesis. I encountered numerous hurdles throughout my research, but his support gave me the clarity needed to continue to push forward, thank you. I would like to thank Dr. Sina Salsabili, who laid the foundation in which my research could be built on. I would also like to thank Dr. Bernard Thébaud and Marissa Lithopoulos, from the University of Ottawa, Ottawa Hospital Research Institute, and Children's Hospital of Eastern Ontario Research Institute, for providing the histopathology WSIs of mouse lung tissue.

Words cannot express how grateful and fortunate I am for my mother and father, who sacrificed so much to allow for their children to have opportunities such as these. I thank my mother, father, grandmother, brothers, and sisters for their unwavering support. Thank you to my loved ones who have allowed me to push for my goals, no matter the difficulty or challenges that arose as a result.

Table of Contents

| | |
|---|-------------|
| Abstract..... | ii |
| Acknowledgements | iii |
| Table of Contents | iv |
| List of Tables | viii |
| List of Figures..... | ix |
| List of Appendices..... | xv |
| List of Abbreviation..... | xvi |
| Chapter 1: Introduction | 1 |
| 1.1 Thesis Objectives..... | 6 |
| 1.2 Thesis Contributions..... | 6 |
| 1.3 Thesis Outline..... | 10 |
| Chapter 2: Background & Literature Review | 12 |
| 2.1 Lung Anatomy..... | 12 |
| 2.2 Histopathology Images of the Lung | 13 |
| 2.3 Emphysematous Changes to Lung Parenchyma..... | 15 |
| 2.4 Conventional MLI Quantification | 17 |
| 2.4.1 Indirect MLI Method (Intersection Counting) | 17 |
| 2.4.2 Direct MLI Method (Chord Measurements) | 19 |
| 2.5 Current Methods for Automated MLI Scoring..... | 23 |
| Chapter 3: Modular Pipeline for Automated MLI Quantification | 28 |
| 3.1 Overview of the Automated System Pipeline..... | 28 |
| 3.2 Image Preprocessing..... | 29 |
| 3.3 Semantic Segmentation | 29 |
| 3.4 Field-of-view Extraction | 33 |

| | | |
|---|--|-----------|
| 3.5 | Field-of-View Screening | 34 |
| 3.6 | MLI Estimation | 36 |
| 3.7 | MLI Score..... | 36 |
| Chapter 4: Indirect Method for MLI Scoring..... | | 38 |
| 4.1 | Intersection Counting Module | 38 |
| 4.2 | Methodology..... | 40 |
| 4.2.1 | Histopathology Dataset of Lung Tissue | 40 |
| 4.2.2 | MLI Scoring: Automated System versus a Single Rater..... | 40 |
| 4.2.3 | Analysis..... | 41 |
| 4.3 | Results | 42 |
| 4.4 | Analysis and Discussion..... | 43 |
| 4.4.1 | Mann-Whitney U Test..... | 44 |
| 4.4.2 | Inter- and Intra-Rater Variability | 46 |
| 4.4.3 | Processing Time: Automated System vs. Human Rater..... | 48 |
| 4.5 | Conclusions | 49 |
| Chapter 5: Implementation of Direct MLI Scoring..... | | 50 |
| 5.1 | Direct MLI Module | 51 |
| 5.2 | Biases between Direct and Indirect MLI Scoring..... | 53 |
| 5.2.1 | Septa Bias..... | 54 |
| 5.2.2 | Partial Chord Bias | 56 |
| 5.2.3 | Combined Septa Bias and Partial Chord Bias | 58 |
| 5.3 | Methodology..... | 59 |
| 5.3.1 | Histopathology Dataset of Lung Tissue | 59 |
| 5.3.2 | MLI Scoring: Indirect Method versus Direct Method..... | 60 |
| 5.3.3 | Analysis..... | 61 |
| 5.4 | Results | 61 |

| | | |
|---|---|-----------|
| 5.4.1 | Direct MLI Scoring | 61 |
| 5.4.2 | Indirect versus Direct Method for MLI scoring | 62 |
| 5.4.3 | Septa Bias, Partial Chord Bias, and Combined Septa Bias + Partial Chord Bias | 64 |
| 5.5 | Analyses and Discussion | 66 |
| 5.6 | Conclusion..... | 68 |
| Chapter 6: Effect of Number of Accepted FOV Images on MLI Score..... | | 70 |
| 6.1 | Number of Accepted FOV Images | 70 |
| 6.2 | Methodology..... | 71 |
| 6.2.1 | Histopathology Dataset of Lung Tissue | 71 |
| 6.2.2 | MLI Scoring..... | 71 |
| 6.3 | Results | 72 |
| 6.4 | Analyses and Discussion | 75 |
| 6.5 | Conclusion..... | 76 |
| Chapter 7: Effect of Guideline Length on MLI Score..... | | 77 |
| 7.1 | Guideline Length | 77 |
| 7.2 | Methodology..... | 80 |
| 7.2.1 | Histopathology Dataset of Lung Tissue | 80 |
| 7.2.2 | MLI Scoring..... | 80 |
| 7.3 | Results | 81 |
| 7.4 | Analyses and Discussion | 88 |
| 7.5 | Conclusion..... | 91 |
| Chapter 8: Conclusions and Future Work | | 92 |
| 8.1 | Conclusions | 92 |
| 8.2 | Study Limitations | 94 |
| 8.3 | Future Work..... | 95 |

| | | |
|-------------------------|---|------------|
| 8.3.1 | Development of a User-Friendly Graphical-User-Interface..... | 95 |
| 8.3.2 | Investigation into analysis of WSIs of lung tissue for different species | 95 |
| 8.3.3 | Addition of other methods for evaluating air space size and emphysematous changes | 96 |
| 8.3.4 | Investigation into the effects of orientation and positioning of the superimposed guideline..... | 96 |
| Appendix A | | 98 |
| References | | 100 |

List of Tables

| | |
|---|----|
| Table 4.1 RMSD between human rater and automated system for each WSI group. | 43 |
| Table 4.2 Mann-Whitney U Test Results: Automated System vs. Human Rater ($\alpha=0.05$). | 45 |
| Table 5.1 RMSD between the indirect and direct MLI scores. | 63 |
| Table 5.2 Average MLI Score per Group for Direct (MLIdir), Indirect (MLIind), <i>Septa Bias</i> (MLIindS), <i>Partial Chord Bias</i> (MLIindE), and <i>Septa Bias + Partial Chord Bias</i> (MLIindSE) MLI Scores..... | 65 |
| Table 5.3 Mann-Whitney U Test Results: Direct (MLI dir) Method for MLI Score ($\alpha=0.05$)..... | 67 |

List of Figures

| | |
|--|----|
| Figure 1.1 Example Field-of-View (FOV) images for MLI scoring using a) indirect method, where intersections (red) between the horizontal guideline (blue) and the alveolar septa are measured, and b) direct method, where distances between consecutive alveolar septa (green) along the length of the guideline are measured. | 4 |
| Figure 2.1 Anatomy of the lungs with major structures labeled [adapted from 22]. | 13 |
| Figure 2.2 Histopathological image of lung tissue with major biological structures labeled. | 14 |
| Figure 2.3 Histopathology image of a) healthy mouse lung tissue, b) diseased lung tissue with the destruction of lung parenchyma. | 16 |
| Figure 2.4 Example FOV images for indirect MLI scoring with a) rejected FOV image due to intersection with pleural space, b) rejected FOV image due to intersection with a bronchus, c) rejected FOV image due to intersection with a blood vessel, d) accepted FOV image with 5 intersections recorded (marked in red). | 18 |
| Figure 2.5 Example FOV images for Direct MLI scoring with a) a single chord measurement (shown in green), b) seven chord measurements. | 20 |
| Figure 2.6 a) Partial Chords (blue), chords (green) and intersection crossings (red) in Direct MLI Scoring, b) partial chords discarded, c) partial chords extended. | 21 |
| Figure 2.7 Example FOV image for direct MLI scoring using a solid test line localized to the left, followed by a dashed guard line. When the test line intersects and alveolar septa, the distance (chord) to the next wall is measured. If a partial chord falls within the guard line, it is extended until it reaches an alveolar septa. | 22 |
| Figure 3.1 Block Diagram of the proposed automated system pipeline. | 28 |

| | |
|---|----|
| Figure 3.2 b) Semantic segmentation versus c) instance segmentation of nuclei from a) histopathology image. | 30 |
| Figure 3.3 Semantic Segmentation binary masks. 1) Alveoli Border Wall, 2) Alveoli Lumen, 3) Bronchi, 4) Blood Vessels, and 5) Background. | 32 |
| Figure 3.4 Example CSV file content containing all data recorded for MLI scoring (first 5 entries). <i>Original Image Name</i> stores the name of the analyzed WSI. <i>FOV Name</i> is used to identify each extracted FOV image. The pixel locations of the two opposite corners of each FOV image are stored under the <i>FOV corner x1, x2, y1, y2</i> headings. <i>ACC/REJ</i> Heading stores whether the FOV image was accepted or rejected and why. <i># Crossings</i> stores the number of intersection crossings for each accepted FOV image. <i># ACC Images</i> stores the total number of accepted FOV images. <i>Total Crossings</i> stores the sum of all the intersection crossings from all the accepted FOV images. <i>MLI Score</i> stores the MLI score. | 34 |
| Figure 3.5 FOV image results from <i>Field-of-view Screening</i> module. a) rejected pleural space, b) rejected bronchi, c) rejected blood vessel and d) accepted FOV images. | 35 |
| Figure 4.1 Block Diagram of the proposed automated system pipeline with indirect MLI method implemented through the intersection counting module. | 38 |
| Figure 4.2 Results from indirect mli module with a) one intersection and b) two intersections. | 39 |
| Figure 4.3 MLI scores from the proposed automated system versus the MLI score (MLI_{ind}) of the human rater. | 42 |
| Figure 4.4 Bland-Altman plot for MLI score (MLI_{ind}) of automated system versus the human rater. | 43 |

| | |
|---|----|
| Figure 4.5 Calculated MLI scores (MLI_{ind}) from the proposed automated system and three different human raters. | 47 |
| Figure 4.6 Bland-Altman plot for the MLI score (MLI_{ind}) of the automated system versus the average score of the 3 human raters. | 48 |
| Figure 5.1 Block Diagram of the proposed automated system pipeline with direct MLI (MLI_{dir}) method implemented. | 50 |
| Figure 5.2 a) The guideline fits entirely inside an alveoli (left) and therefore both ends are extended until they reach an alveolar border wall (right). b) Both ends fall inside an alveoli (left) and are therefore extended until they reach an alveolar border wall (right). c) Left side of guideline falls inside an alveolar septa while the right side falls within an alveoli (left) and therefore the left side of the guideline is not extended and the right side of the guideline is extended until it reaches the edge of the FOV image, since it did not reach an alveolar septa (right). d) Left edge of the guideline falls within an alveoli (left) but is not extended because doing so would result in contact with a bronchi (right). | 52 |
| Figure 5.3 Example CSV file content containing all data recorded for MLI scoring using the direct method (first 5 entries). <i>Original Image Name</i> stores the name of the analyzed WSI. <i>FOV Name</i> is used to identify each extracted FOV image. The pixel locations of the two opposite corners of each FOV image are stored under the <i>FOV corner x1, x2, y1, y2</i> headings. <i>ACC/REJ</i> heading stores whether the FOV image was accepted or rejected and why. <i>Image Chord</i> stores which FOV image is being analyzed, with the chord measurement stored directly to the right under the <i>Chord Measurements</i> header. <i># ACC Images</i> stores the total number of accepted FOV images. <i># Chords</i> stores the total number of chords measured from all the accepted FOV images. <i>Average Chord Length</i> | |

| | |
|--|----|
| stores the average chord length (L_m), which is the MLI score calculated through the direct method..... | 53 |
| Figure 5.4 <i>Septa Bias</i> FOV image, where G_0 represents the original length of the guideline while septa widths are represented by S_1 , S_2 , and S_3 | 55 |
| Figure 5.5 <i>Partial Chord Bias</i> FOV image, where G_E represents the total length of the new guideline, which is the sum of chord measurements (L) and septa widths (S). | 57 |
| Figure 5.6 Combined <i>Septa Bias + Partial Chord Bias</i> FOV image, where G_{SE} represents the total length of the guideline composed of the sum of chord measurements ($L(k)$). ... | 58 |
| Figure 5.7 Example chord length distribution obtained from automated direct MLI (MLI_{dir}) scoring. | 62 |
| Figure 5.8 MLI score calculated by the automated system using the indirect (MLI_{ind}) and direct (MLI_{dir}) MLI modules. | 63 |
| Figure 5.9 Bland-Altman plot for MLI score for indirect (MLI_{ind}) and direct (MLI_{dir}) methods. | 64 |
| Figure 5.10 MLI Score using indirect (MLI_{ind}), direct (MLI_{dir}), <i>Septa Bias</i> (MLI_{indS}), <i>Partial Chord Bias</i> (MLI_{indE}), and <i>Septa Bias + Partial Chord Bias</i> (MLI_{indSE}) equations. | 65 |
| Figure 6.1 Average mean MLI score and standard error for varying number of accepted FOV images for RA group..... | 73 |
| Figure 6.2 Average mean MLI score and standard error for varying number of accepted FOV images for RA + LPS group | 73 |
| Figure 6.3 Average mean MLI score and standard error for varying number of accepted FOV images for O_2 group..... | 74 |

| | |
|---|----|
| Figure 6.4 Average mean MLI score and standard error for varying number of accepted FOV images for O ₂ + LPS group..... | 74 |
| Figure 7.1 Example FOV image for a) indirect MLI scoring with large guideline, b) indirect MLI scoring with small guideline, c) direct MLI scoring with large guideline, d) direct MLI scoring with small guideline..... | 79 |
| Figure 7.2 Average indirect MLI (MLI _{ind}) score (solid line) and average number of intersections (dashed line) vs. guideline length (in μm). | 81 |
| Figure 7.3 Average Direct MLI (MLI _{dir}) Score (solid line) and Average Number of Chords (dashed line) vs. Guideline Length (in μm). | 82 |
| Figure 7.4 Number of accepted FOV images as a function of guideline length..... | 83 |
| Figure 7.5 Average indirect MLI (MLI _{ind}) score per WSI group vs. guideline length (μm) at fixed number of accepted FOV images per WSI (400) | 84 |
| Figure 7.6 Average direct MLI (MLI _{dir}) score per WSI group vs. guideline length (μm) at fixed number of accepted FOV images per WSI (400) | 85 |
| Figure 7.7 Average indirect MLI score (MLI _{ind}) and direct MLI score (MLI _{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for RA + LPS WSI Group | 86 |
| Figure 7.8 Average indirect MLI score (MLI _{ind}) and direct MLI score (MLI _{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for RA WSI Group | 86 |
| Figure 7.9 Average indirect MLI score (MLI _{ind}) and direct MLI score (MLI _{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for O ₂ WSI Group | 87 |

| | |
|---|----|
| Figure 7.10 Average indirect MLI score (MLI_{ind}) and direct MLI score (MLI_{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for $O_2 +$ LPS WSI Group | 87 |
|---|----|

List of Appendices

| | |
|-----------------|----|
| Appendix A..... | 98 |
|-----------------|----|

List of Abbreviation

| | |
|------------|--|
| ARDS | Acute Respiratory Distress Syndrome |
| CNN | Convolutional Neural Network |
| COPD | Chronic Obstructive Pulmonary Disease |
| DSC | Dice Similarity Coefficient |
| FOV | Field-of-View |
| MLI | Mean Linear Intercept |
| MSC | Mesenchymal Stem Cell |
| PA | Pixel-wise Accuracy |
| PPV | Positive Predictive Value |
| SARS-CoV-2 | Sever Acute Respiratory Syndrome Coronavirus 2 |
| WSI | Whole Slide Image |

Chapter 1: Introduction

Pulmonary emphysema is a common and preventable chronic lung condition that is characterized by the gradual damage and deterioration of the alveoli in the lung. It is part of a group of lung diseases known as chronic obstructive pulmonary disease (COPD). Respiratory symptoms include difficulty breathing, coughing, chest tightness, and phlegm production. Smoking tobacco is the leading cause of pulmonary emphysema, but environmental factors such as air pollutants, fumes, and chemicals can also cause the disease to develop [1]. COPD is the third leading cause of death worldwide, responsible for approximately 3.23 million deaths in 2019 alone [1]. There are no effective treatments to cure emphysema, with current interventions aimed at reducing the progression of the disease or managing current symptoms. In the latter stages of the disease, airways can become largely obstructed while lung parenchyma are severely destroyed, limiting the effectiveness of any currently available interventions [1], [2].

Respiratory illnesses associated with severe acute respiratory syndrome coronavirus 2 (*i.e.*, SARS-CoV-2), which is the virus that causes COVID-19, have recently become a large focus in research [67]-[69]. Lung complications such as pneumonia, acute respiratory distress syndrome (ARDS), and sepsis can occur as a direct result of SARS-CoV-2, with long-term complications associated with the disease still not fully understood [3]. With the prevalence of COPD amongst the general population and the widespread transmission of SARS-CoV-2, the relationship between COVID-19 and patients with pre-existing COPD could result in severe complications. Zhao *et al.* performed a meta-analysis on the impact

of COPD on COVID-19 severity, finding that the risk of developing severe COVID-19 in patients with pre-existing COPD was four times higher than in patients without [4].

Novel treatments for COPD, such as mesenchymal stem cell therapy, are in the early stages of research [5]-[7]. In one study, bone marrow-derived cells were shown to contribute to the regeneration and repair of lung tissue of cigarette-exposed mice [8]. More recently, mesenchymal stem cell therapy has begun to be researched as a method for treating COVID-19. Infection with COVID-19 may cause an immune overreaction within the body, resulting in large amounts of inflammatory factors and overproduction of immune cells and cytokines, known as a cytokine storm [9]. Recently, Chen *et al.* [10] researched the effectiveness of mesenchymal stem cells (MSCs) in treating ARDS induced by epidemic influenza (H7N9) infection. 17 patients with H7N9-induced ARDS were treated with allogeneic menstrual-blood-derived MSCs and acted as the experimental group, while 44 patients with H7N9-induced ARDS were left as a control group. Those treated through MSC transplantation had a significantly lower mortality rate (17.6 %) compared to that of the control group (54.5%) [10]. The immunomodulatory and anti-inflammatory properties of MSC treatment for respiratory diseases were confirmed in 17 completed clinical studies, with approximately 70 ongoing trials registered through the U.S. National Library of Medicine (ClinicalTrials.gov) [10]. These findings warrant further investigation into the potential of stem cell therapies as treatments for COPD and COVID-19. Although many investigations are still in their infancy, this field of research is profoundly growing. Researchers often use animal models in their investigation, as they can mimic chronic lung diseases common to humans and display the associated lung injury phenotype.

Researchers must be able to accurately and objectively quantify lung structure when researching potential therapies in order to assess treatment effectiveness and disease progression. Qualitative observations of pulmonary emphysema are subjective and potentially biased methods for determining the impact of the disease, and therefore quantitative metrics are needed. The mean linear intercept (MLI) is a popular and widespread metric for quantifying lung structure in histopathological images of lungs [11]. The MLI is a measure of the mean free distance of the air spaces, and has been thoroughly applied to investigate emphysematous changes to lung parenchyma and smoke-induced lung injury [12], [13]. It is commonly used to quantify the enlargement of the airspaces, which may result from structural damage caused by emphysema; an increase in airspace size is reflected as an increase in MLI [14]. Morphometric parameters, such as the Destructive Index [65] and Section Assessment [66], are also used. The Destructive Index is a measure of alveolar septa damage and emphysema, representing the percentage of destroyed space as a fraction of the total alveolar and duct space [65]. Section Assessment involves the grading of different sections of all lung lobes by comparison with pictures from [66] to grade emphysema. While parameters such as Destructive Index and Section Assessment are used, the most heavily utilized technique is the MLI, appearing most frequently in the relevant literature [64].

Current methods to quantify the MLI typically involve the use of microscopic analysis of histopathological images by an expert. In this technique, hundreds of sub-images are extracted from a histopathology lung whole slide image (WSI). In this thesis, these sub-images are referred to as field-of-view (FOV) images. A horizontal guideline is

superimposed onto each FOV image. MLI can be measured indirectly, which involves counting the number of times the guideline intersects with an alveolar septa (Figure 1.1.a). The MLI is calculated as the total length of the guidelines, divided by the number of intersections, from all FOV images. The MLI can also be measured directly, by measuring chord lengths, which are the distances along the guideline between consecutive alveolar septa (Figure 1.1.b) [15]. The MLI score is an estimate of the average chord length.

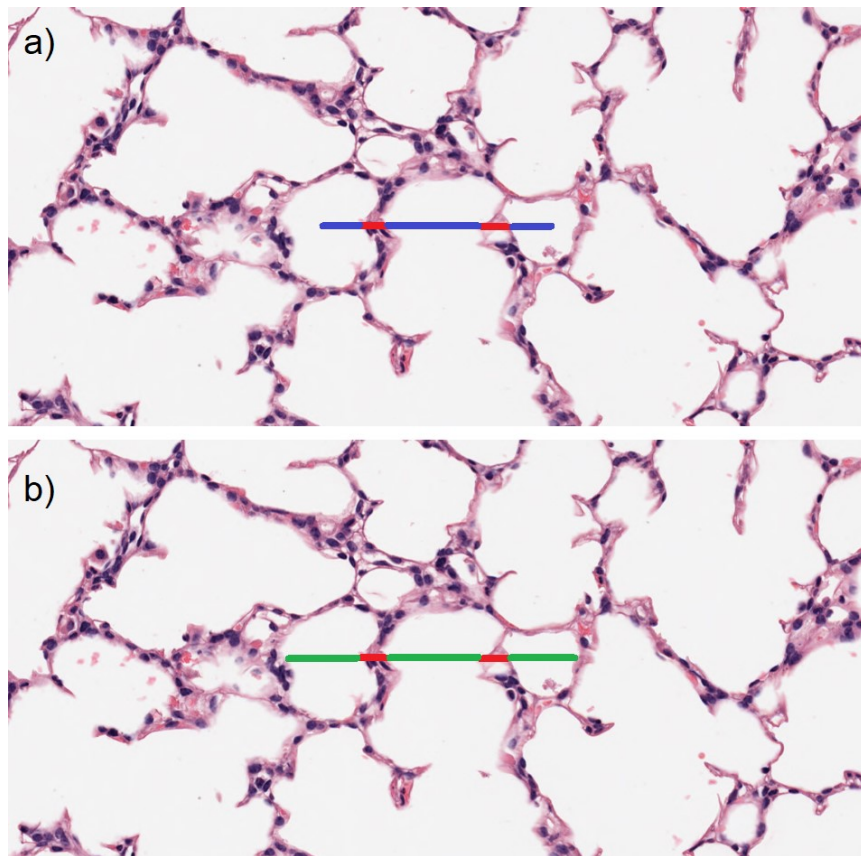


Figure 1.1 Example Field-of-View (FOV) images for MLI scoring using a) indirect method, where intersections (red) between the horizontal guideline (blue) and the alveolar septa are measured, and b) direct method, where distances between consecutive alveolar septa (green) along the length of the guideline are measured.

The MLI score is extremely inefficient when performed manually. Human raters must first determine if a FOV image must be rejected and indicate the reason (ex. rejection due to contact with a blood vessel). For the remaining accepted FOV images, the human rater must then count and record the number of intersection crossings. The time required to evaluate a single lung WSI is typically on the order of hours, with [37] noting an approximate assessment time of 10 hours per WSI. Furthermore, such methods are subjective due to a lack of visual gold standards when encountering different biological structures found within the lung tissue, resulting in inter- and intra-rater variability [11], [15]. For example, a human rater may include an additional intersection crossing when analyzing a FOV image while a different human rater may not. Repeated differences in intersection crossings will then result in inter-rater variability in the final MLI scores. Similarly, if a human rater repeats MLI scoring on the same FOV images, they may have differences in the number of intersections counted for the same FOV image, resulting in intra-rater variability in the final MLI scores. Therefore, current methods for MLI quantification are time-consuming, inefficient and susceptible to biases.

Efforts have been made to automate histopathology image analysis to address the shortcomings of manual analysis techniques [16]-[18]. However, factors such as stain variation, artifacts, and biological structure variability have proven to be challenging hurdles in developing robust, accurate and fully automated solutions. Accurate segmentation of biological structures found in lung histopathology images is one of the main obstacles in automating the MLI scoring process, which is challenging given the high

degree of variability within these structures. Recent advances in biomedical image segmentation, specifically through the use of deep learning methods, have led to significant improvements in segmentation accuracy and performance [19]. These advancements have enabled the potential to fully automate the MLI scoring process, which would lead to a significant increase in efficiency when analyzing histopathology images of lung tissue and allow researchers to quantify results for a much higher volume of specimens.

1.1 Thesis Objectives

The overall long-term objective of this thesis is the development of a fully automated system that can accurately calculate the MLI score, given a digitized histopathology image of lung tissue.

The objectives of the thesis are as follows:

1. Develop a modular system/pipeline to fully automate the MLI scoring process using digitized histopathology images of mouse lung tissue
2. Compare indirect and direct MLI scoring techniques
3. Investigate the effect of the number of accepted FOV images on the MLI score
4. Investigate the effect of guideline length on the MLI score

1.2 Thesis Contributions

The work presented in the thesis resulted in four main contributions:

1. Development of a modular pipeline that fully automates indirect MLI scoring given digitized histopathology images of mouse lung tissue and performance evaluation against human raters

Previous work to fully automate the indirect scoring of the mean linear intercept was demonstrated in 2021 by Salsabili *et al.* [37]. The author noted segmentation errors due to hand-crafted features, which may have contributed to errors in the MLI score. Furthermore, the design was highly specialized, meaning changes to even small segments of the model would require modification to the entirety of the pipeline. In this thesis, a fully automated pipeline was implemented using a modular approach, allowing for individual modules of the pipeline to be modified or enhanced without the need to make changes to other areas of the pipeline. The fully automated pipeline was evaluated on 20 mouse lung WSIs against the average manual assessment of three human rater, with a root-mean-squared deviation of 4.71 ± 2.16 in the MLI score while displaying a high correlation ($r=0.9955$).

2. Development of a modular pipeline that fully automates direct MLI scoring given digitized histopathology images of mouse lung tissue and comparison against indirect techniques

Previous work to fully automate the direct scoring of the mean linear intercept was described by Sallon *et al.* [39] through colour thresholding and custom software. The accuracy of their method, however, was poor, with differences in MLI score as great as 40 μm for multiple WSIs. Their method also failed to differentiate between the biological structures found within the WSI, meaning the measurements were not a true representation of the alveolar and ductal air spaces. In this thesis, a fully

automated pipeline was implemented using a modular approach, allowing for individual modules of the pipeline to be modified or enhanced without the need to make changes to other areas of the pipeline. The automated system utilized a semantic segmentation module, which allowed for differentiation between the alveolar/ductile air spaces and other biological structures such as blood vessels and bronchi.

Indirect measurement of the MLI score is done through intersection counting between the guideline and alveolar septa, while the direct measurement is performed by measuring distances between consecutive alveolar septa along a guideline. Although both indirect and direct measurements look to calculate MLI score, the results differ with the direct MLI scoring being consistently lower. In this thesis, the alveolar septa width and fixed guideline length are shown to be the major biases responsible for the differences between the two methods. The results indicate that the alveolar septa width and fixed guideline length each contributed to approximately half the difference between the indirect and direct MLI scores.

3. Investigation into the effects of varying the number of field-of-view images used for MLI scoring

The MLI scoring process involves analyzing field-of-view images extracted from the original WSI and measuring exact distances or intersections for the direct and indirect methods, respectively. One parameter that heavily influences the resulting MLI score is the number of accepted FOV images used in the calculation. While a

minimum of 300 FOV images are recommended for MLI scoring [37], [38], the impact of varying this parameter on actual WSIs of lung tissue has not been explored.

Results demonstrate that as the number of accepted FOV images increases, the standard error in the MLI score decreases. The average standard error when using only 50 accepted FOV images to calculate the MLI score was as high as 4.31 and 7.26 for direct and indirect methods, respectively. The average standard error when using 300 accepted FOV images to calculate the MLI score was as high as 1.72 and 2.81 for direct and indirect methods, respectively. Therefore the recommended number of 300 accepted FOV images from previous literature appears to be a reasonable, however a further increase in the number of accepted FOV images still leads to a decrease in standard error. The average standard error when using 1000 accepted FOV images to calculate the MLI score was as high as 1.38 and 0.83 for direct and indirect methods, respectively. The automated system allows the user to set the stride length between FOV images, with little cost to set them to higher values, therefore allowing for a large number of accepted FOV images. This allows for a reduction in the resulting standard error and is recommended.

4. Investigation into the effects of varying the guideline length used for MLI scoring

Guideline length is also an important parameter when performing MLI scoring. The effect of varying guideline length on the resulting MLI score has not been thoroughly analyzed in literature.

Results demonstrate that indirect MLI scoring decreases with increasing guideline length, with a relationship that is similar to an exponential decay. For direct MLI scoring, results show that guideline length has little impact, with the relationship being relatively flat. The average standard error when using a guideline length of 19 μm to calculate the MLI score was as high as 1.98 and 52.37 for direct and indirect methods, respectively. The average standard error when using a guideline length of 155.34 μm was as high as 1.45 and 2.31 for direct and indirect methods, respectively. The average standard error when using a guideline length of 349 μm was as high as 1.12 and 1.68 for direct and indirect methods, respectively. The automated system allows the user to set the guideline length used for MLI scoring, with little cost to set them to higher values. This allows for a reduction in the resulting standard error and is recommended.

1.3 Thesis Outline

The thesis is divided into 8 chapters. Chapter 1 introduces the research purpose, states the research objectives in relation to the clinical problem, and summarizes the research contributions. Chapter 2 presents the relevant background knowledge and a literature review detailing previous works on automating the MLI scoring process. Chapter 3 provides an overview of the pipeline for the automated system. Chapter 4 describes the

implementation of the pipeline utilizing the indirect MLI scoring technique and presents the evaluation of the automated system against manual MLI scoring by human raters. Chapter 5 describes the implementation of the pipeline utilizing the direct MLI scoring technique and presents an analysis of the differences between the indirect and direct MLI scoring techniques. Chapter 6 presents an analysis of how the number of accepted FOV images affects the MLI score. Chapter 7 presents an analysis of how guideline length affects the MLI score. Chapter 8 presents the conclusions and areas for future work.

Chapter 2: Background & Literature Review

This chapter provides the background on lung structure, histopathological images, effects of emphysematous changes to lung parenchyma, and the most common methods for MLI scoring. This chapter also provides a literature review of previous works on automating MLI scoring.

2.1 Lung Anatomy

The lungs are the fundamental organs of the respiratory system, whose primary responsibility is the efficient facilitation of gas exchange between the environment and the bloodstream. Figure 2.1 shows the anatomy of the lungs with major structures labeled. In its simplest form, the lungs provide the necessary passageway that allows oxygen to diffuse into the bloodstream while allowing for carbon dioxide to diffuse out. This exchange of gases occurs in millions of tiny air sacs called alveoli, which are enveloped in a network of thinly walled blood vessels called capillaries. A thin layer of cells, on average 1 micron (1/10000 of a centimeter) in thickness, lines the alveoli and surrounding capillaries; this barrier allows for efficient movement of gases between the alveoli and capillaries. Alveoli are separated from each other by thin border walls known as septa. Air that is inhaled through the mouth or nose travels through the trachea and into the right and left bronchus, which travels into smaller passageways called bronchioles. Air passes through these bronchioles and into alveolar ducts, which finally connect to the network of capillaries.

The outer layer of the lungs, blood vessels, and bronchi are covered by a serous membrane known as visceral pleura [20], [21].

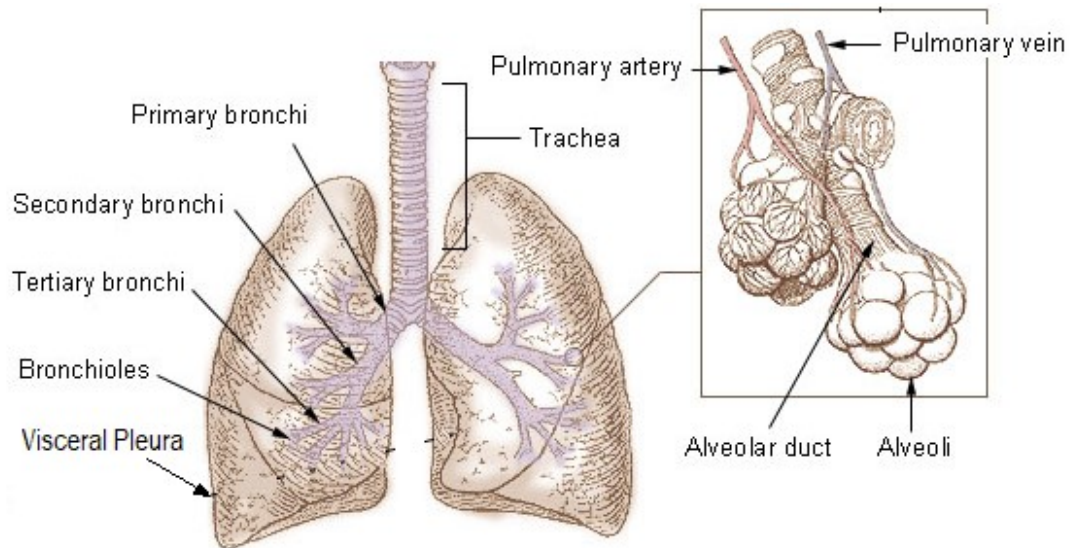


Figure 2.1 Anatomy of the lungs with major structures labeled [adapted from 22].

2.2 Histopathology Images of the Lung

Lung histopathology is the microscopic analysis and diagnosis of diseased tissues of the lung. Lung tissue is typically removed from the body and is immediately fixated, which is a process used to permanently preserve tissue structure; fixation kills pathogenic microorganisms, hardens tissue, stops cell metabolism, and prevents enzymatic degradation of cells and tissues. After fixation, the specimen is washed, dehydrated and enclosed in an embedding medium, which is typically paraffin. Embedding preserves the tissue morphology and gives structural support to the tissue during sectioning. When the specimen has cooled and hardened, it is mounted on a specialized machine called a microtome, which allows for extremely thin slices of the specimen to be cut. These thin slices of lung tissue are then stained, typically with solutions of hematoxylin and eosin

(H&E), which greatly enhances the contrast between different biological structures in the tissue, such as bronchi, blood vessels, and septa. The specimens can then be mounted on glass slides for microscopic analysis [23], [24]. Figure 2.2 displays a histopathology image of lung tissue (described in Section 2.2), with major biological structures identified. Blood vessels, which carry blood for gas exchange throughout the lungs, can be typically identified by the red blood cells found inside the structure (red blood cells may not always be present). The nuclei of cells, such as those found in the walls of blood vessels and bronchi, are typically dyed purple as a result of the hematoxylin staining [25]. The space found outside of the lung structure is known as the pleural space, which is the cavity that exists between the outside of the lungs and the inside of the chest wall [21].

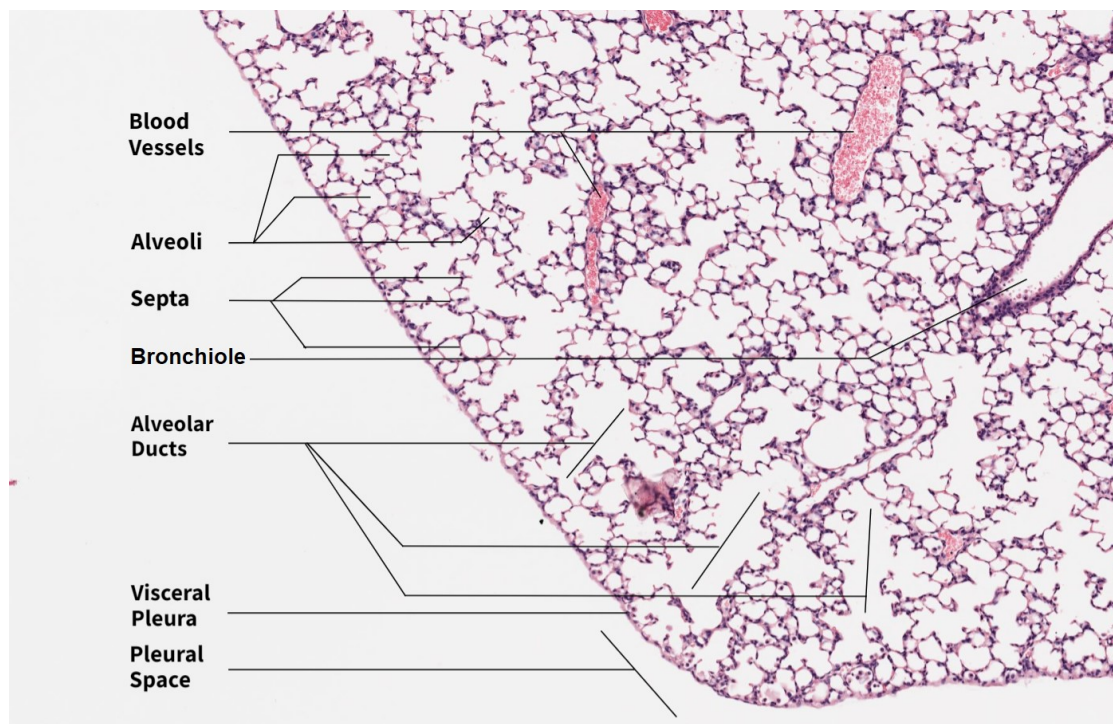


Figure 2.2 Histopathological image of lung tissue with major biological structures labeled.

2.3 Emphysematous Changes to Lung Parenchyma

Long term exposure to air pollutants, fumes, chemicals and smoking tobacco are major factors in the development of emphysematous changes to lung parenchyma. The development and progression of the disease represent a major challenge to public health due to its severity and frequency, as current health-care systems may not be able to cope with the negative long-term projection trends [26]. In healthy lungs, the homeostatic balance between protease and anti-protease enzymes regulates the regeneration, repair, and long-term health of lung parenchyma. Smoking tobacco and environmental factors cause an inflammatory response in the airways and lung parenchyma, and dysregulation between protease and anti-protease enzymes [27]. The imbalance of these enzymes results in the destruction of alveolar septa, alveolar ducts, and respiratory bronchioles. This loss in alveolar septa results in larger and less frequent alveoli, which reduces the total surface area available for gas exchange. There is also a reduction in elastic recoil of the lung (*i.e.*, the ability for the lungs to rebound after being stretched during inhalation) and a loss of radial traction of the airways (*i.e.*, the force exerted by the lung parenchyma to keep the airways open), which results in the retention of excess air within the lungs during respiration [28]. Figure 2.3 illustrates the destruction of lung parenchyma associated with pulmonary emphysema by contrasting it against healthy lung tissue. The much larger and less frequent alveoli from the diseased lung tissue are evident through visual examination alone. However, researchers require quantitative measurements to objectively assess disease progression or treatment effectiveness.

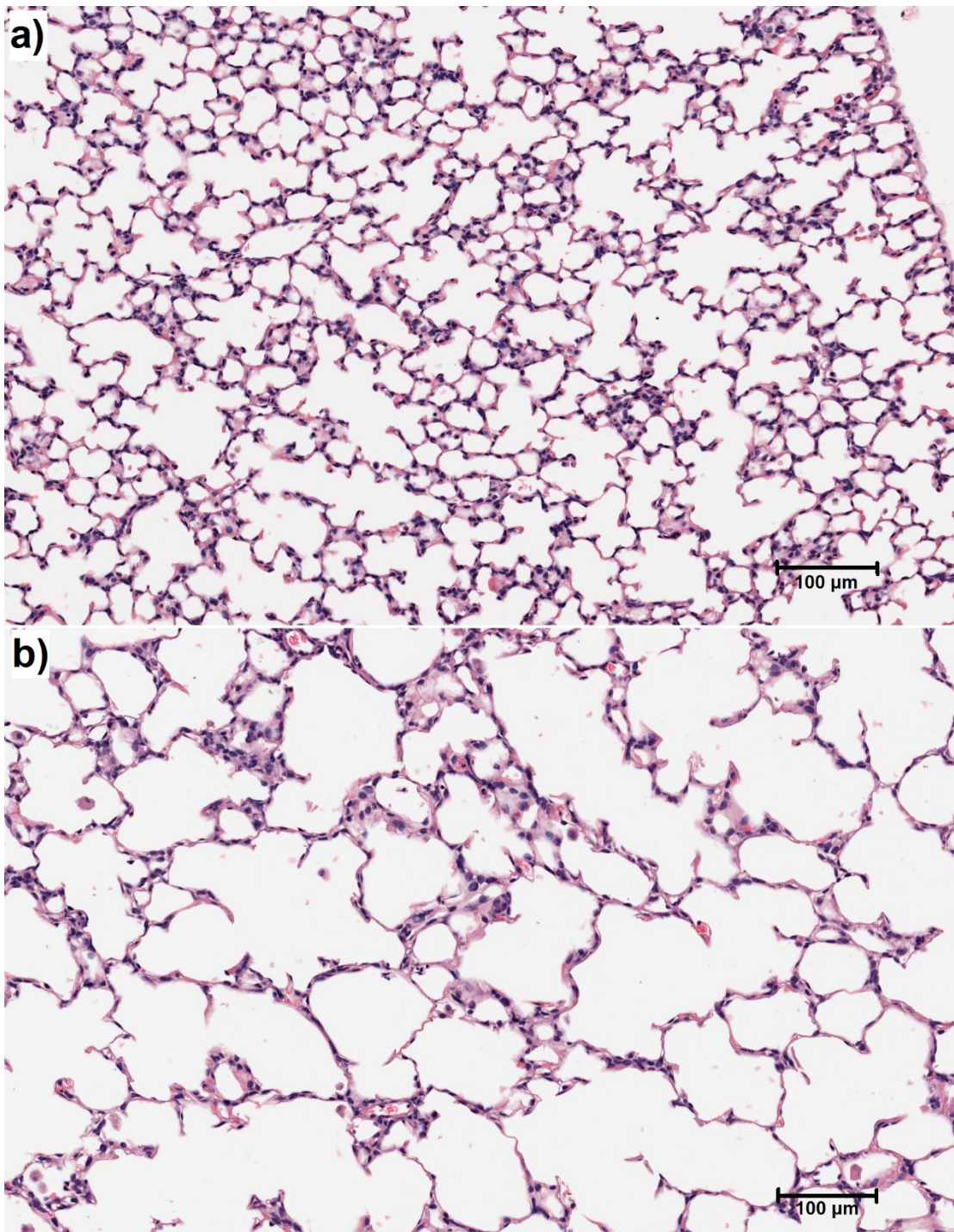


Figure 2.3 Histopathology image of a) healthy mouse lung tissue, b) diseased lung tissue with the destruction of lung parenchyma.

2.4 Conventional MLI Quantification

The mean linear intercept (MLI) is a widely used metric for quantifying emphysematous changes to lung parenchyma and has been frequently applied to investigate such changes in experimental pathology and studies on human lung disease [29], [30], [31]. The MLI is most accurately described as a measure of the mean free distance of the air spaces, including the alveoli and alveolar ducts, and has been shown to increase with disease progression [11], [30]. Although one of the most popular measures, the MLI is significantly dependent on lung fixation, as the calculation is severely impacted by changes in lung volume [32]. Collapse and disruption to lung structure caused by errors in the fixation process can cause lung volume changes, alveolar wall thickening and fixation artifacts, which could lead to inaccurate or misleading measurements of the MLI score [33]. With proper precautions and accurate fixation processes, which can be comfortably achieved with currently available techniques, meaningful interpretation of the MLI score can be attained [30], [34]. The MLI can be calculated indirectly, through intersection counting, and directly through measurements of the distances between alveolar septa.

2.4.1 Indirect MLI Method (Intersection Counting)

The indirect method for determining the MLI score is often conducted in a semi-automated process, where field-of-view (FOV) images are automatically presented to a human rater for analysis. The FOV images are sub-images, generated from pre-defined grid points using microscopic software (e.g., MetaMorph Software version 7.8, Molecular Devices, LLC). A horizontal guideline of fixed length is superimposed in the center of each FOV image, which is used by the rater for MLI quantification. [35], [36]. The rater must first identify if

a FOV image can be used for MLI calculation (accepted FOV image) or if it must be excluded (rejected FOV image). A FOV image is rejected if any part of the guideline intersects with the pleural space (i.e., outside the lung space; Figure. 2.4a), a bronchus (Figure. 2.4b), or a blood vessel (Figure. 2.4c). The remaining FOV images are accepted and used to calculate the MLI score. For each accepted FOV image, the rater counts the number of intersections, which is the number of times the guideline fully crosses over an alveolar border wall (Figure 2.4d).

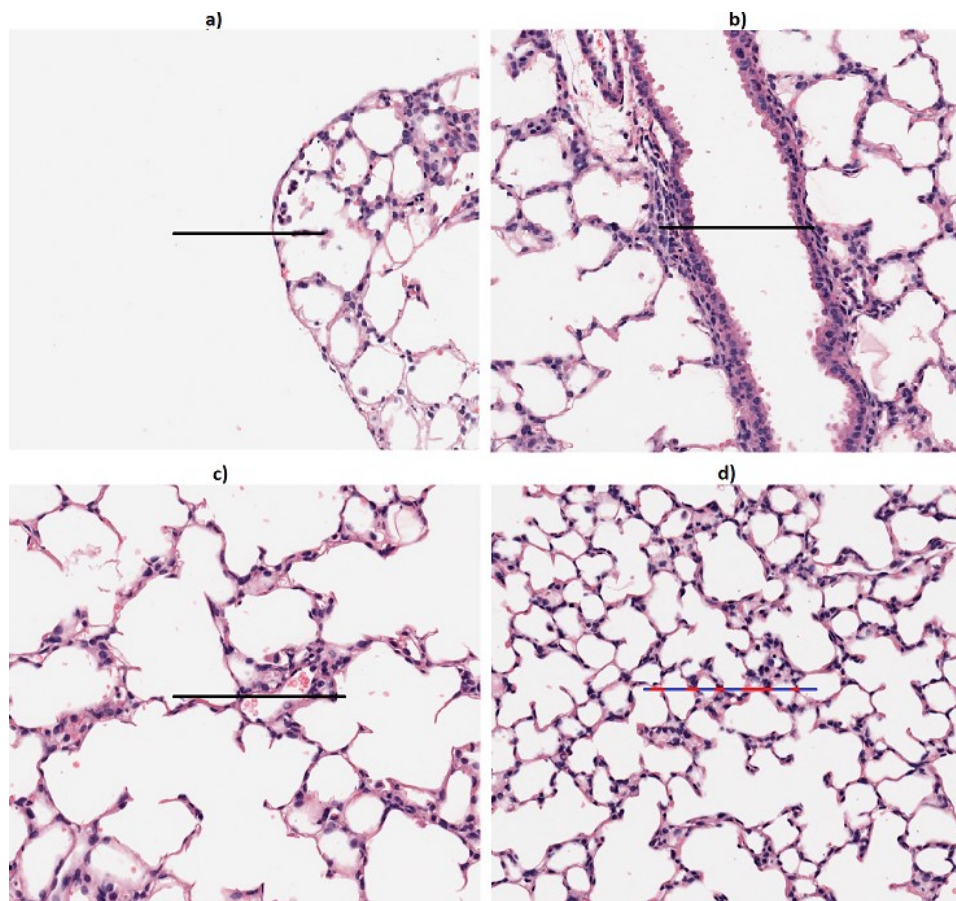


Figure 2.4 Example FOV images for indirect MLI scoring with a) rejected FOV image due to intersection with pleural space, b) rejected FOV image due to intersection with a bronchus, c) rejected FOV image due to intersection with a blood vessel, d) accepted FOV image with 5 intersections recorded (marked in red).

The MLI score (in μm) for the indirect method can then be calculated (Equation 2.1)

$$MLI_{ind} = \frac{\sum_{i=1}^N G(i)}{\sum_{i=1}^N C(i)} \quad (2.1)$$

where N is the number of FOV images that were accepted, G is the length of the guideline superimposed on the i^{th} FOV image (in μm), and $C(i)$ is the number of intersection crossings for the i^{th} FOV image. In the conventional method for indirect MLI scoring, the guideline has a fixed length (e.g., a length of $155.34 \mu\text{m}$ was used in [37]). A minimum number of accepted FOV images is often recommended (e.g., at least 300 accepted FOV images) [37], [38]. With a fixed guideline length, equation 2.1 simplifies to equation 2.2

$$MLI_{ind} = \frac{N \times G_0}{\sum_{i=1}^N C(i)} \quad (2.2)$$

where N is the number of FOV images that were accepted, G_0 is the fixed length of the guideline superimposed on each FOV image (in μm), and $C(i)$ is the number of intersection crossings for the i^{th} FOV image.

2.4.2 Direct MLI Method (Chord Measurements)

Similar to the indirect method, the direct method for MLI scoring is often conducted in a semi-automated process, where FOV images are automatically presented to a human rater for analysis. The FOV images are generated from pre-defined grid points using microscopic software, where each FOV image has a superimposed horizontal guideline used for MLI quantification [38]. The rater must identify if a FOV image can be used for MLI calculation (accepted FOV image) or if it must be excluded (rejected FOV image). Identical to the indirect method, a FOV image is rejected if any part of the guideline

intersects with the pleural space (i.e., outside the lung space), a bronchus, or a blood vessel. The remaining FOV images are accepted and used to calculate the MLI score. Unlike the indirect method, where the rater counts the number of intersections between the guideline and the alveolar septa, the rater instead measures chord lengths; a chord is a section of the guideline spanning consecutive alveolar septa (Figure 2.5).

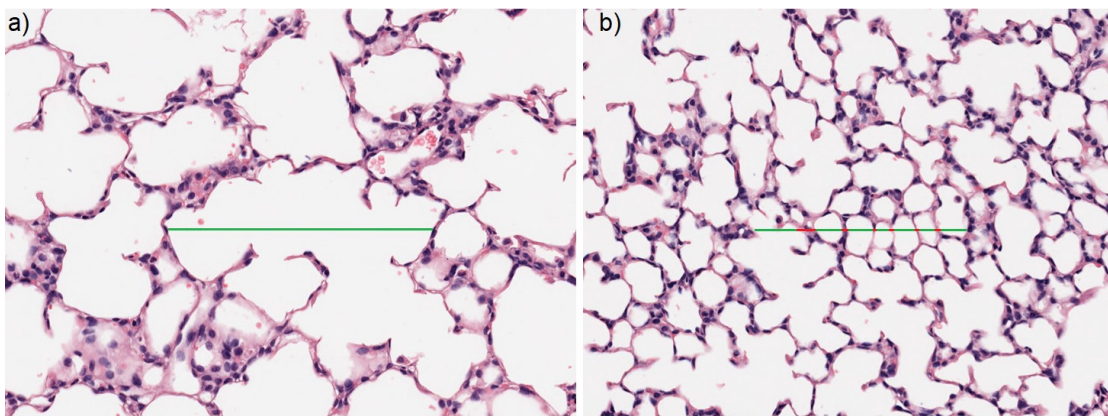


Figure 2.5 Example FOV images for Direct MLI scoring with a) a single chord measurement (shown in green), b) seven chord measurements.

When a guideline of fixed length is superimposed on a FOV image, the ends can often fall within alveoli (Figure 2.6a). These sections of the guideline are referred to as partial chords. Partial chords can be discarded (Figure 2.6b), as to not negatively skew the measurements with smaller chords. Partial chords can also be extended (Figure 2.6c) until the ends of the guideline reach another alveolar septa.

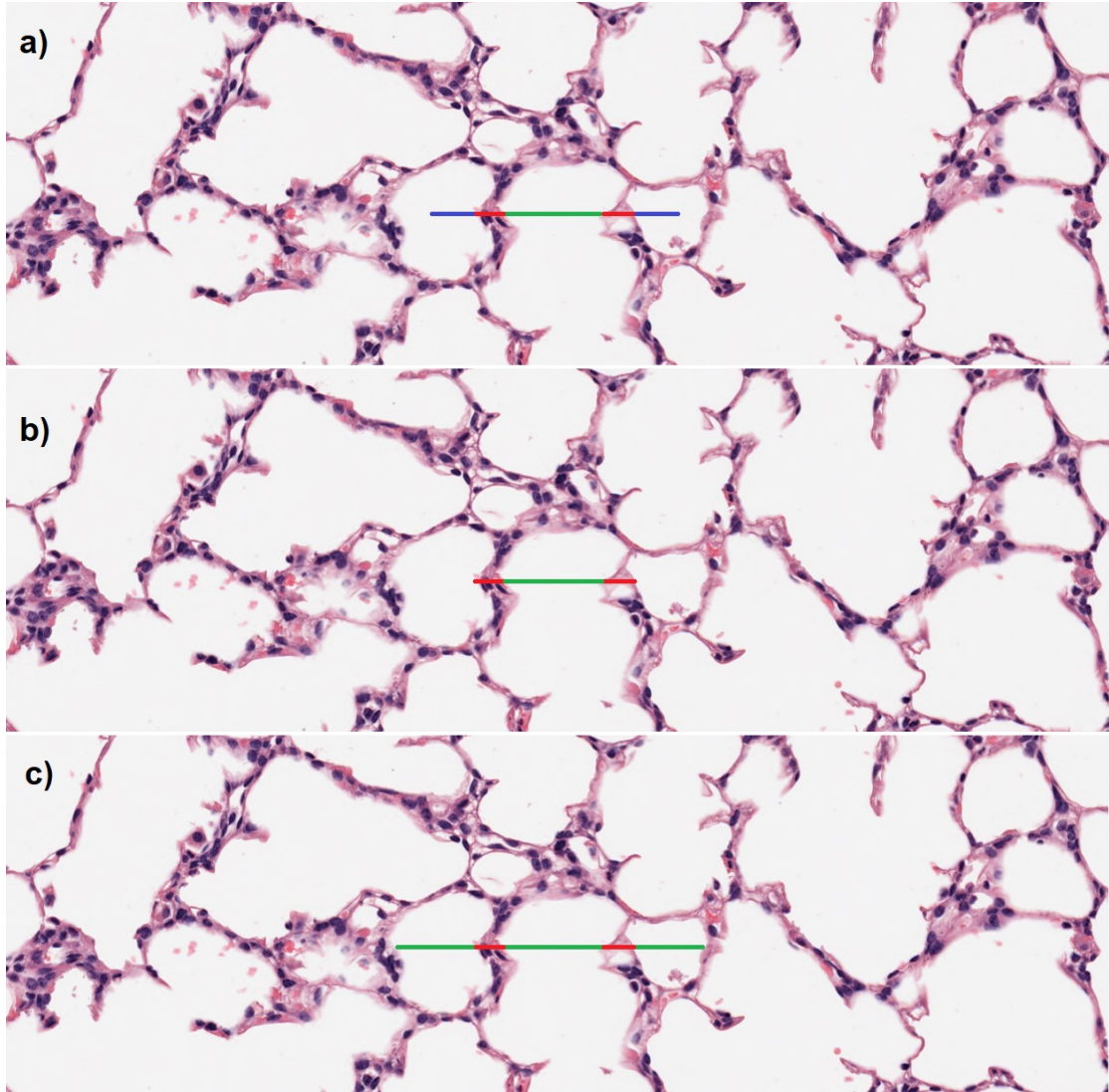


Figure 2.6 a) Partial Chords (blue), chords (green) and intersection crossings (red) in Direct MLI Scoring, b) partial chords discarded, c) partial chords extended.

The technique of extending partial chords is often conducted in direct MLI scoring [38], [41], [42]. In [38], a solid guideline, known as the test line, is fixed towards the left side of a FOV image, where chords are measured along the guideline (from left to right) from one alveolar septa to the next (Figure 2.7). A longer dashed guideline, known as the guard line, is superimposed to the right of test line. In this approach, partial chords that fall within the

guard line are extended until they reach an alveolar septa. This approach ensures that longer chords are not undersampled [38].

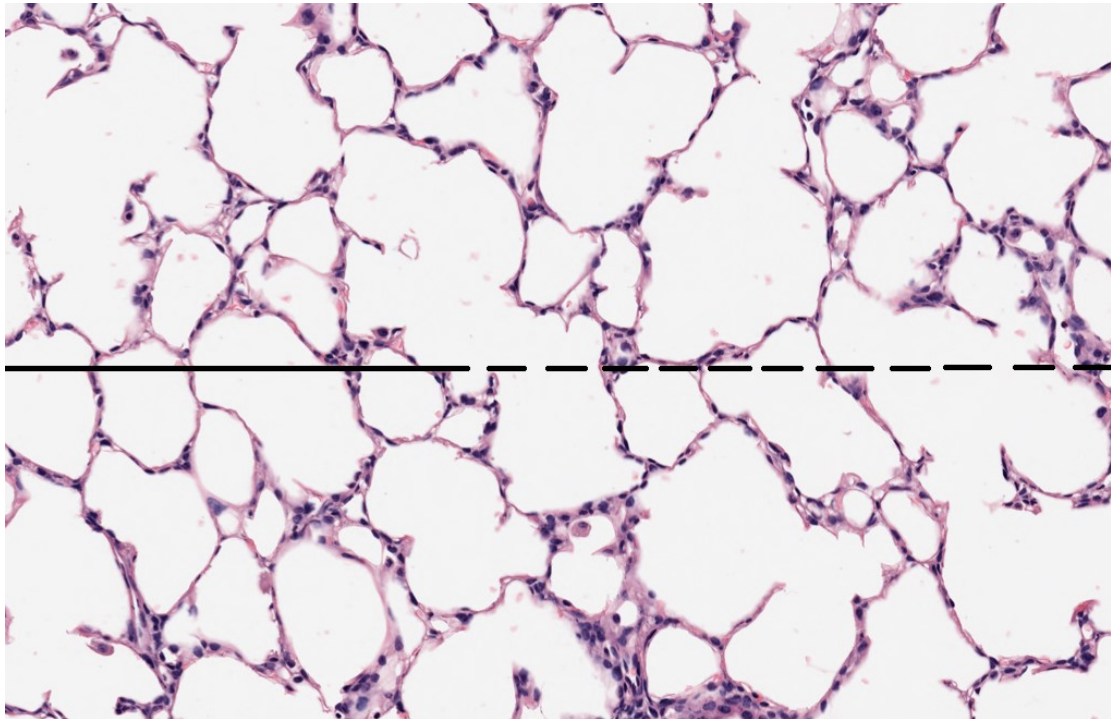


Figure 2.7 Example FOV image for direct MLI scoring using a solid test line localized to the left, followed by a dashed guard line. When the test line intersects an alveolar septa, the distance (chord) to the next wall is measured. If a partial chord falls within the guard line, it is extended until it reaches an alveolar septa.

In this paper, the guideline is superimposed in the center of the FOV image, and partial chords on either end of the guideline are extended until they reach an alveolar septa. The resulting MLI score is calculated by finding the mean chord length (L_m). By measuring individual chords (L) directly, the rater is able to obtain the variance of the airspace measures, which provides a better understanding of the alveolar surface area of the lungs.

Although the direct method provides more information compared to the indirect method, it is more time-consuming and labor-intensive [38].

The MLI score for the direct method can then be calculated (Equation 2.3)

$$MLI_{dir} = L_m = \frac{\sum_{i=1}^N \sum_{k=1}^{K(i)} L_i(k)}{\sum_{i=1}^N K(i)} \quad (2.3)$$

where N is the total number of accepted FOV images, $K(i)$ is the number of chords in the i^{th} FOV image, and $L_i(k)$ is the length (in μm) of k^{th} chord in the i^{th} FOV image.

2.5 Current Methods for Automated MLI Scoring

The most popular methods for MLI scoring are based on the semi-automated assessment of histopathology images, which are inefficient, expensive, and susceptible to biases [29], [30], [35], [36]. Efforts have been made toward fully automating the MLI scoring process to address the shortcomings associated with semi-automated techniques. One of the most challenging hurdles in developing a fully automated and robust solution to calculate the MLI score is the accurate segmentation of biological structures found within histopathology images of lung tissue. Identifying these structures is challenging, given their high degree of variability in shape and size, while failing to properly segment these structures during fully automated processes can result in underestimating the MLI score [39].

Published in 2006, Parameswaran *et al.* [11] developed an automated method for measuring linear intercepts (chords) from digitized histopathology images of lung tissue and calculated the mean chord length, L_m , from their distribution. Their method used an

image-thresholding approach to separate tissue from alveolar air space and superimposed a uniform grid of horizontal and vertical lines to measure chords. The authors determined the threshold value manually or with computer assistance, which is inefficient and subjective for each WSI. Furthermore, image thresholding is susceptible to a variety of errors, since even optimal thresholds may result in segmentation errors of the tissue structures. Their approach also did not segment the various biological structures (i.e., bronchi and blood vessels) from the alveoli and alveolar ducts, likely leading to underestimation of the MLI score [39]. Their method was only applied and tested on a small sub-set of FOV images, specifically curated by the authors, rather than with whole slide images (WSIs). Furthermore, their results were not compared against the manual assessment by human raters, meaning that there was no quantitative evidence of its accuracy. Such an approach would therefore have difficulty generalizing to the various WSIs that may be encountered and is an inadequate method for automated MLI scoring.

Rieger-Fackeldey *et al.* [40] studied the effects of hyperoxia development by analyzing histopathology images of lung tissue from newborn mice. They utilized digital image analysis software (Image Pro Plus version 4.0), a custom written macro, and colour segmentation to determine alveolar morphological characteristics. Using only colour variation to segment alveoli from the remaining tissue can lead to segmentation errors, especially given the colour variability between and within WSIs, due to differences in the staining process. Furthermore, similarity in colour between alveoli, blood vessels, and bronchi could further increase segmentation errors. The authors provided no technical details regarding the design of their proposed automated algorithms nor the accuracy of

their method compared to manual assessment by human raters. Therefore this method would likely have difficulty generalizing to various WSIs and would also be an inadequate method for automated MLI scoring.

In 2015, Sallon *et al.* [39] proposed a more promising automated method to calculate the MLI score in WSIs of mouse lung tissue. They used a thresholding approach combined with an analysis of closed areas based on size to identify alveolar structures. Furthermore, they assessed the accuracy of their algorithm against the manual assessment of two human raters. The authors' proposed method, however, was applied to as little as 10 FOV images per WSI. In contrast, manual/semi-automated methods typically require hundreds of FOV images per WSI for MLI scoring [37], [38]. Furthermore, the accuracy of their proposed method was questionable, as the results showed statistically significant differences between the two human raters and the automated method for multiple WSIs. Finally, their proposed method failed to identify or differentiate the main biological structures (*i.e.*, bronchi, blood vessels, alveoli) within the lungs, which could lead to underestimation of the MLI score, as noted by the authors.

Recently, Salsabili *et al.* [37] proposed a method to segment the various biological structures in lung images to facilitate automated MLI scoring. Segmentation was achieved through a combination of colour thresholding, morphological operations, and region growing techniques, and created segmentation masks to distinguish these structures from one another. In segmenting the various biological structures, their method achieved a mean accuracy, Dice coefficient, and Hausdorff distance of 98.34%, 98.22%, and 109.68 μm .

These segmentation masks were used to calculate the MLI score via the indirect method. When evaluated against the manual assessment by a human rater across 10 mouse lung images, results showed a mean difference and standard deviation in MLI score of 2.33 ± 4.13 . While the results were good, their techniques for identifying the various biological structures were based on hand-crafted features selected by the authors, which they noted caused some segmentation errors. Furthermore, their method performed MLI scoring through the indirect method alone. Direct measurements of chord lengths can display the variance of the airspace measures and, therefore, a better understanding of alveolar surface volume. While the direct method is more tedious and time-consuming than the indirect method with manual assessment, it is relatively simple in automated assessments once the images are segmented.

In 2022, Salsabili *et al.* [49] introduced a novel two-stage multiresolution segmentation algorithm capable of segmenting complex biological structures found in mouse lung tissue. Four different CNNs were used to extract contextual information at varying magnifications (i.e., 2.5 \times , 5 \times , 10 \times , and 20 \times), with a single CNN to aggregate the extracted information and generate the final segmentation masks. The proposed method had mean \pm standard deviation of 0.971 ± 0.024 , 0.873 ± 0.033 , and 0.833 ± 0.028 for the pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV), respectively. The high segmentation accuracy of the proposed method provides an opportunity to develop a usable and accurate system for MLI scoring, and is utilized in the fully automated pipeline described in this thesis.

To the author's knowledge, this work is the first system that utilizes a modular pipeline to fully automate the MLI scoring process given digitized histopathology images. The modularity of the pipeline allows for efficient enhancements or modifications to individual modules without requiring changes to other areas of the pipeline. The automated system's simplistic design requires only that the user provides a digitized WSI of the lung tissue as input, where all further steps are taken care of by the pipeline, producing detailed information about the analysis of the respective WSI and the final MLI score as output. The proposed automated system is also the only method that allows for both indirect and direct MLI scoring, providing the user with the flexibility to utilize which technique they prefer. The proposed automated system is also used to provide an in-depth analysis of different parameters such as guideline length and number of accepted FOV images in order to assess their impact on the resulting MLI score. Due to the inefficiencies of semi-automated methods (e.g., MLI scoring can take hours per image, inter- and intra-rater variability), such in-depth analyses have not been performed previously.

Chapter 3: Modular Pipeline for Automated MLI Quantification

This chapter presents an overview of the designed automated pipeline used to calculate the MLI score. The modularity of the pipeline allows for efficient enhancements or modifications to individual modules without requiring changes to other areas of the pipeline.

3.1 Overview of the Automated System Pipeline

Figure 3.1 shows a block diagram of the proposed automated system, with each module enclosed and arrows showing the flow of the pipeline. The input to the pipeline is a digitized WSI of lung tissue and the output is the MLI score. There are five processing modules in the pipeline: 1) Image Preprocessing, 2) Semantic Segmentation, 3) Field-of-View Extraction, 4) Field-of-View Screening, and 5) MLI Estimation. Each module is described in the following subsections.

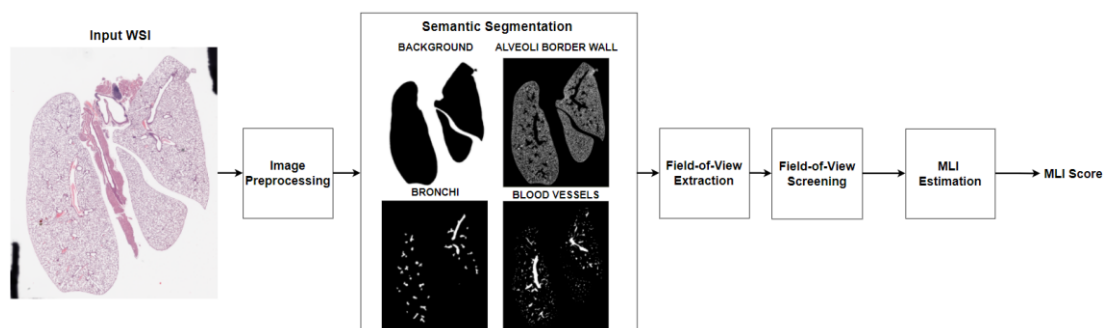


Figure 3.1 Block Diagram of the proposed automated system pipeline.

3.2 Image Preprocessing

Image preprocessing is used to improve image quality, suppress undesired distortions or artifacts and enhance key features which may be necessary depending on the particular application. This module provides an opportunity to enhance the image quality to an acceptable standard or limit factors that may negatively impact subsequent steps in the pipeline. One important image preprocessing step is colour normalization, which mitigates colour variations that can arise in histological images due to a number of factors such as histochemical staining time and histology stain quantity. These variations in colour may negatively impact segmentation performance and subsequent MLI score. This implementation of the *Image Preprocessing* module colour normalizes the incoming WSI through a colour mapping technique [43] to mitigate any colour variations. This colour normalization technique was successfully used as an image preprocessing step for image segmentation [44]-[46]

3.3 Semantic Segmentation

One of the largest hurdles in automating the MLI scoring process is the inability to accurately identify the various biological structures found within histopathological images of lungs. Variations in these structures, such as in shape, size and colour make it difficult to differentiate between them in automated methods. However, for accurate MLI scoring, one must be able to differentiate between intersection crossings of the guideline with the alveolar septa versus intersection crossings with a bronchus or blood vessel, and when the guideline is within the lung structure or in the pleural space. Similarly, one must be able to record measurements between alveolar septa (chords), rather than distances between

alveolar septa and bronchi, blood vessels or pleura. Failure to accurately differentiate between these biological structures has been shown to result in an underestimation of the MLI score [39].

In the *Semantic Segmentation* module, the objective is to accurately segment the various biological structures of the lung. In semantic segmentation, each pixel within the WSI is labeled to a specific class, with all pixels within a single class being labeled as one entity (Figure 3.2b). This differs from instance segmentation, where each object within these classes is labeled individually (Figure 3.2c). For the fully automated system, identification of individual alveoli, blood vessels and other biological structures is not required; only differentiation between the different classes is needed for MLI scoring. The outputs of the *Semantic Segmentation* module are binary masks from one of five classes: 1) Alveoli Border Wall, 2) Alveoli Lumen, 3) Bronchi, 4) Blood Vessels, and 5) Background, which are all other pixels and include the pleural space and tissue that are not of interest.

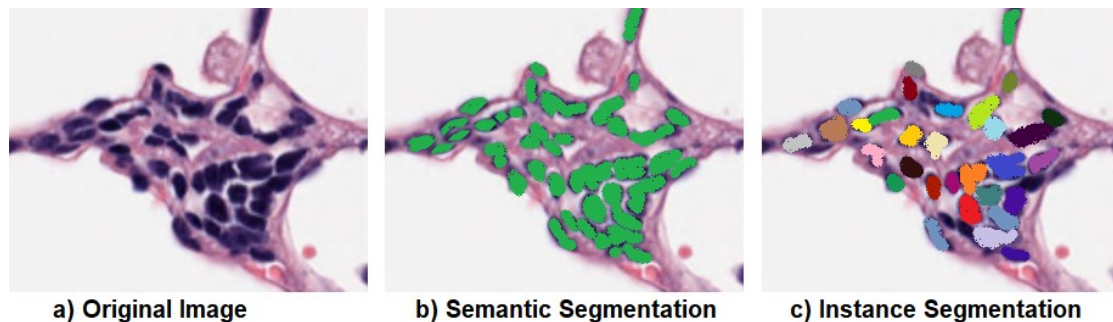


Figure 3.2 b) Semantic segmentation versus c) instance segmentation of nuclei from a) histopathology image.

In previous work, hand-crafted features based on characteristics such as colour, area, shape, and textural features were used to perform semantic segmentation [37]. More recently, deep learning methods, specifically convolutional neural networks (CNN), have demonstrated effectiveness in the semantic segmentation for other histopathological WSIs [47], [48]. Histopathological WSIs are large, high-resolution images (*e.g.*, a single WSI used in this thesis can be $80,000 \times 80,000$ pixels in size) that cannot be directly inputted to CNNs. Therefore, WSIs are often divided into small patches (*e.g.*, size 256×256 pixels) that can then be processed by a CNN. These patches are individually segmented and later combined to create the final segmented binary masks. While smaller patches may be easier to process by a CNN in terms of computational efficiency, they may not provide enough information for accurate segmentation, which can be particularly true for larger biological structures that may be larger than the patch size. A solution is to increase the patch size; however this is at the cost of computational efficiency, which typically leads to increased training times for the model. Alternatively, the original WSI can first be down-sampled prior to patch extraction, resulting in a wider field-of-view (FOV) for a given patch size. However, this approach reduces the resolution of the patch and may affect model performance, especially for smaller biological structures. There is a three-fold tradeoff between field-of-view, spatial resolution, and computational efficiency. This tradeoff is specifically troublesome for histopathology WSIs of lung tissue due to the large size variation of the different biological structures. For example, histopathology WSIs of lung tissue may have blood vessels as small as $24 \mu\text{m}^2$ and bronchi as large as approximately 1.2 mm^2 [49].

Recent work attempts to resolve this three-fold tradeoff through a two-stage, multiresolution semantic segmentation approach [49]. The multiresolution approach analyzes WSIs at varying magnifications, which is a technique that is often performed by pathologists. In the first stage, four CNNs are used to extract contextual information from the input patches at four different magnifications (i.e., magnification of 2.5 \times , 5 \times , 10 \times and 20 \times). Each CNN generates a segmentation heat-map, associated with its particular magnification. In the second stage, a single CNN is used to aggregate the resulting heat-maps from the first stage to generate the final five segmentation masks: 1) Alveoli Border Wall, 2) Alveoli Lumen, 3) Bronchi, 4) Blood Vessels, and 5) Background, as shown in Figure 3.3.

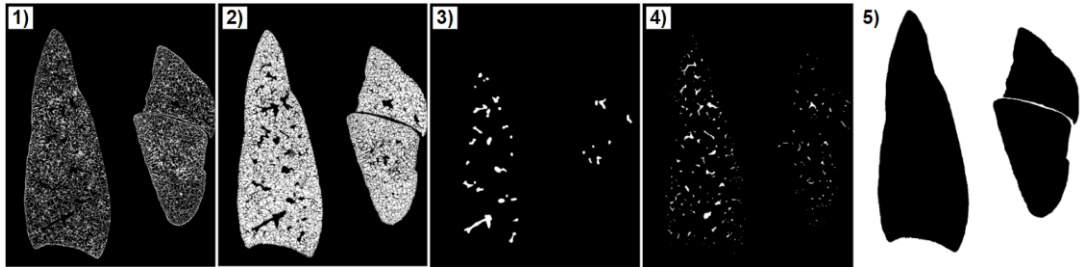


Figure 3.3 Semantic Segmentation binary masks. 1) Alveoli Border Wall, 2) Alveoli Lumen, 3) Bronchi, 4) Blood Vessels, and 5) Background.

The MLI scoring pipeline presented in this thesis uses this multiresolution approach [49]. The modularity of the pipeline allows for new segmentation algorithms to be easily incorporated as they become available, benefiting from advanced methods which may be more efficient or accurate. Furthermore, the proposed pipeline allows for manual intervention, where a user can manually refine the masks before they are utilized in

subsequent stages (e.g., to resolve gross segmentation errors resulting from image artifacts).

3.4 Field-of-view Extraction

In the *Field-of-View Extraction* module, multiple FOV images are extracted from the original WSI and corresponding semantic segmentation masks. This module of the pipeline mimics the conventional semi-automated approach, where FOV images are presented to a rater for MLI scoring. FOV images are extracted using a sliding window approach. The user specifies the size of the FOV images (width \times height) and stride length (x and y) between each FOV image. As an example, for FOV images of size 800×600 pixels, there would be no overlap with a stride length of 800 and 600, 50% overlap with a stride length of 400 and 300, and maximal overlap with a stride length of 1 and 1. By manipulating these parameters, the user can increase the number of FOV images used in the MLI scoring process. The pixel locations of the two opposite corners of each FOV image, with respect to the original WSI ([1, 1] = top left corner of the WSI), are recorded and saved as a CSV file, as shown in Figure 3.4.

| Original Image Name | FOV Name | FOV corner x1 | FOV corner x2 | FOV corner y1 | FOV corner y2 | ACC/REJ | # Crossings | # ACC Images | Total Crossings | MLI Score |
|---------------------|----------|---------------|---------------|---------------|---------------|---------|-------------|--------------|-----------------|-------------|
| O2 + LPS 3 | img_1 | 1 | 1388 | 1 | 1072 | REJ_BG | NaN | 1758 | 3026 | 90.24708048 |
| | img_2 | 348 | 1735 | 1 | 1072 | REJ_BG | NaN | | | |
| | img_3 | 695 | 2082 | 1 | 1072 | REJ_BG | NaN | | | |
| | img_4 | 1042 | 2429 | 1 | 1072 | REJ_BG | NaN | | | |
| | img_5 | 1389 | 2776 | 1 | 1072 | REJ_BG | NaN | | | |

Figure 3.4 Example CSV file content containing all data recorded for MLI scoring (first 5 entries). *Original Image Name* stores the name of the analyzed WSI. *FOV Name* is used to identify each extracted FOV image. The pixel locations of the two opposite corners of each FOV image are stored under the *FOV corner x1, x2, y1, y2* headings. *ACC/REJ* Heading stores whether the FOV image was accepted or rejected and why. *# Crossings* stores the number of intersection crossings for each accepted FOV image. *# ACC Images* stores the total number of accepted FOV images. *Total Crossings* stores the sum of all the intersection crossings from all the accepted FOV images. *MLI Score* stores the MLI score.

Saving the FOV images extracted from the original WSI to storage is not necessary for MLI scoring, but it could be helpful for manual verification or other analyses. Therefore, the system allows the user to select whether or not to save the extracted FOV images. Furthermore, if the user decides to save the FOV images, they can also choose to save the images with the horizontal guideline superimposed or not. The modularity of the pipeline allows for different FOV image extraction methods to be implemented, such as a random FOV image extraction rather than the current sliding window approach.

3.5 Field-of-View Screening

In the *Field-of-view Screening* module, FOV images are screened to remove undesirable images and only utilize those that are deemed acceptable. This implementation mimics the conventional method that manual raters use when screening FOV images. In this approach,

a horizontal guideline of fixed length is superimposed on the center of the FOV image. In this implementation, the length of the guideline is a parameter that can be set by the user. The *Field-of-view Screening* module analyzes the FOV images based on the locations that were saved in the CSV file during the *Field-of-View Extraction* module.

If the superimposed guideline intersects with the pleural space (Figure 3.5a), a bronchus (Figure 3.5b), or a blood vessel (Figure 3.5c), the image is rejected; otherwise, the image is accepted (Figure 3.5d) and is used to calculate the MLI score.

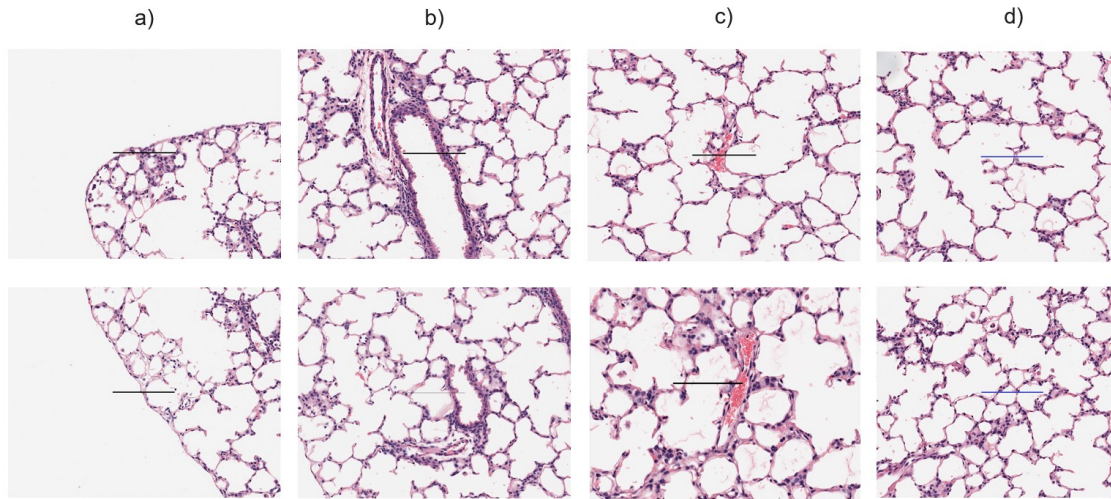


Figure 3.5 FOV image results from *Field-of-view Screening* module. a) rejected pleural space, b) rejected bronchi, c) rejected blood vessel and d) accepted FOV images.

Three binary masks are used to reject FOV images: 1) Bronchi, 2) Blood Vessels, and 3) Background. If the pixels of the superimposed guideline intersect with pixels in at least one of these binary masks, the FOV image is rejected; otherwise the FOV image is accepted. To track which FOV images are accepted and rejected, the module updates the CSV file

(Figure 3.4) with the corresponding result. Under the *ACC/REJ* header, the FOV image will either be accepted (*i.e.*, *ACC*), or rejected with the reason why (*i.e.* *REJ_BG*, *REJ_BR*, or *REJ_BV*).

3.6 MLI Estimation

In the *MLI Estimation* module, all accepted FOV images are used to calculate the MLI score. The *MLI Estimation* module analyzes the accepted FOV images based on the results saved in the CSV file during the *Field-of-View Screening* module. The MLI score can be determined through the indirect method, which relies on intersection counting between the guideline and the alveolar septa, and the direct method, which measures the distances between consecutive alveolar septa along the length of the guideline. Regardless of the method, the Alveolar Border Wall binary mask is used to determine intersection crossings or chord lengths. The modularity of the proposed automated system allows the user to select which method they desire. Depending on which method is implemented, either the number of intersection crossings or the length of the chords measured per FOV image are recorded and saved in the updated CSV file. Chapter 4 and Chapter 5 detail the implementation of the indirect MLI method and the direct MLI method for the *MLI Estimation* module, respectively.

3.7 MLI Score

The final MLI score is calculated using all the accepted FOV images and their respective number of intersection crossings or chord measurements, depending on whether the indirect or direct MLI method was used, respectively. If the indirect method was used, the

MLI score is calculated as per equation (2.1). If the direct method was used, the MLI score is calculated as per equation (2.3). The MLI score is recorded and saved in the updated CSV file.

Chapter 4: Indirect Method for MLI Scoring

The MLI score can be measured indirectly through intersection counting. An intersection occurs when the guideline fully crosses over an alveolar border wall. In this approach, all accepted FOV images are analyzed and the number of intersections for each FOV image is measured and recorded. The pipeline of the proposed automated system using the indirect MLI method is shown in Figure 4.1. This is identical to the generic pipeline shown in Figure 3.1, except for the introduction of the *Intersection Counting* module, replacing the generic *MLI Estimation* module.

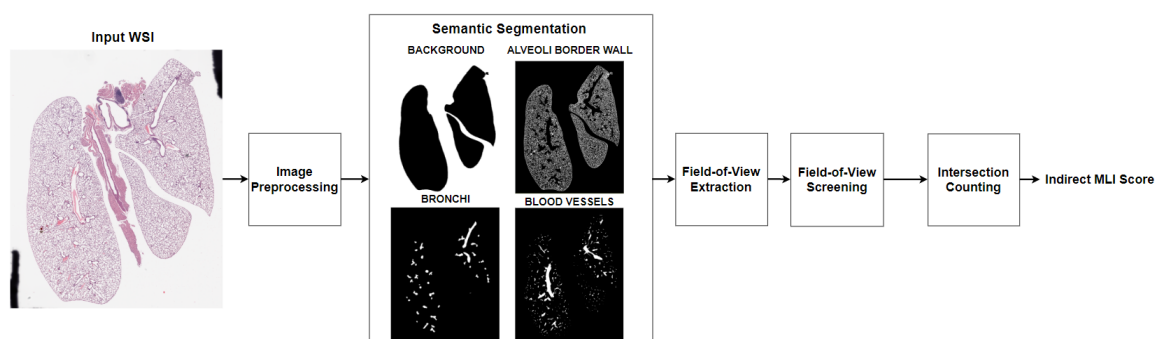


Figure 4.1 Block Diagram of the proposed automated system pipeline with indirect MLI method implemented through the intersection counting module.

In this chapter, the implementation of the *Intersection Counting* module is defined and the proposed automated system is evaluated against the manual assessment by a human rater.

4.1 Intersection Counting Module

In the indirect method for MLI scoring, the number of intersections between the guideline and alveolar septa is recorded for every accepted FOV image. This module mimics the way

that human raters perform indirect MLI scoring in the conventional semi-automated approach. For the *Intersection Counting* module, the same guideline length that was used in the *Field-of-view Screening* stage is used to determine the number of intersections. An intersection is when the guideline fully crosses an alveolar septa. The original guideline length is marked in blue, and intersections between the guideline and alveolar septa are marked in red. Figure 4.2 displays two FOV images that pass through the *Intersection Counting* module. The user can choose to save the FOV images with the guidelines and red-marked intersection crossings; while unnecessary, this option contributes to the transparency of the system, allowing the user to visualize the detected intersections. The intersection counting process is repeated for every accepted FOV image, with the results saved in a CSV file, as shown in Figure 3.4. The MLI score for the indirect method, MLI_{ind} , is calculated using Equation 2.1 from Section 2.4.1, where the sum of all guideline lengths (in μm) is divided by the total number of intersections measured from all accepted FOV.

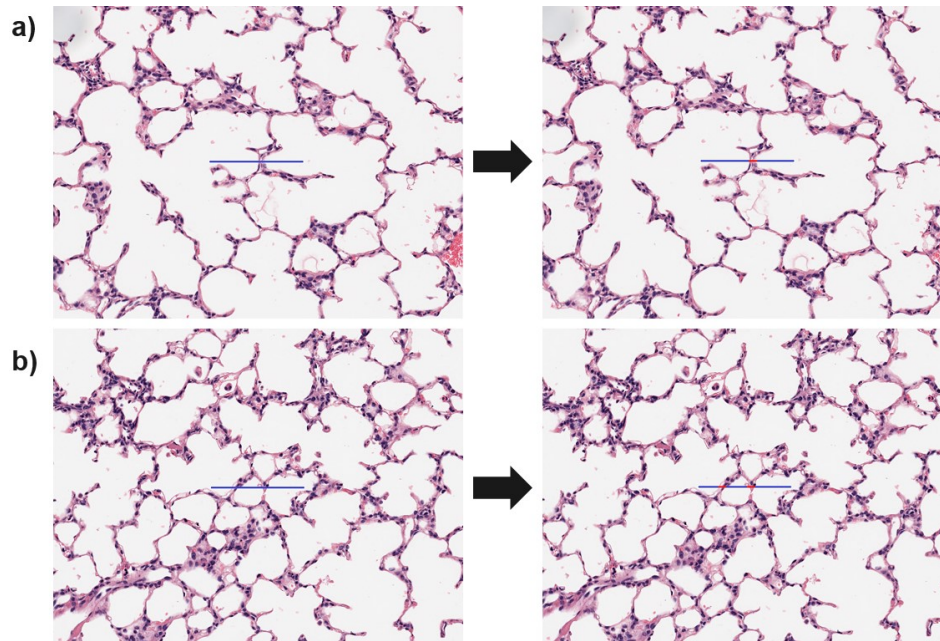


Figure 4.2 Results from indirect mli module with a) one intersection and b) two intersections.

4.2 Methodology

4.2.1 Histopathology Dataset of Lung Tissue

The proposed automated system was tested on a set of 20 WSIs, comparing the MLI score generated from the automated system against the MLI score from a human rater. The dataset consists of 20 histopathology WSIs of mouse lungs, provided by the Sinclair Centre for Regenerative Medicine (Ottawa Hospital Research Institute, Ottawa, ON). The WSIs were prepared in accordance with all policies and protocols set forth by the University of Ottawa Animal Care Committee. All specimens were prepared as detailed in [37]. The 20 mice were from one of four different experimental groups. The first group consisted of 5 healthy mouse lung tissue from mice that were housed in room air (RA group), which acted as the control group. The next group consisted of 5 WSIs of lung tissue from mice that were exposed to a high concentration of oxygen (O_2 group). The next group consisted of 5 WSIs of lung tissue from mice that were housed in room air but were exposed to a high concentration of lipopolysaccharide (RA + LPS group). The final group, consisting of 5 WSIs of lung tissue, was taken from mice that were exposed to both high concentrations of oxygen and lipopolysaccharide (O_2 + LPS group). High concentrations of oxygen and lipopolysaccharide, can lead to lung injury, resulting in typically fewer and larger alveoli within the lung, when compared to those in the RA control group. This should result in a higher MLI score due to the decrease in the number of alveoli and an increase in alveolar size [37].

4.2.2 MLI Scoring: Automated System versus a Single Rater

The proposed automated system described in chapter 3, with the indirect MLI module from chapter 4.1, is used to calculate the MLI score for all 20 WSIs described in chapter 4.2. FOV images of size 1072 (height) \times 1388 (width) pixels were extracted using a sliding window, with a stride length of 268 and 347 in the horizontal and vertical directions, respectively (75% overlap). The number of accepted FOV images per WSI ranged from 782 to 1846. The horizontal guideline length was set to 312 pixels, which corresponds to a length of 155.34 μm for the given scanning resolution (this length has been used for MLI scoring of mouse lungs; e.g., [51], [52])

The MLI score for all 20 WSIs was also calculated in previous works [49] by a human rater who was engaged in clinical research involving MLI scoring of mouse lungs and is considered an expert rater with three years of experience. Semi-automated scoring was performed using Metamorph Software version 7.8, Molecular Devices, LLC. Note that this software used different FOV images than the automated system, but used the same guideline length of 155.34 μm .

4.2.3 Analysis

The MLI scores of the automated were compared against the MLI scores of the human rater using root-mean-squared deviation (RMSD), a Bland Altman analysis [50], and a Pearson correlation coefficient.

4.3 Results

The resulting MLI scores from the automated system and the human rater for the 20 WSIs are shown in Figure 4.3. Table 4.1 summarizes the RMSD between the MLI score of the human rater and the automated system. There was a very strong correlation ($r=0.9931$) between the MLI scores of the human rater and the automated system.

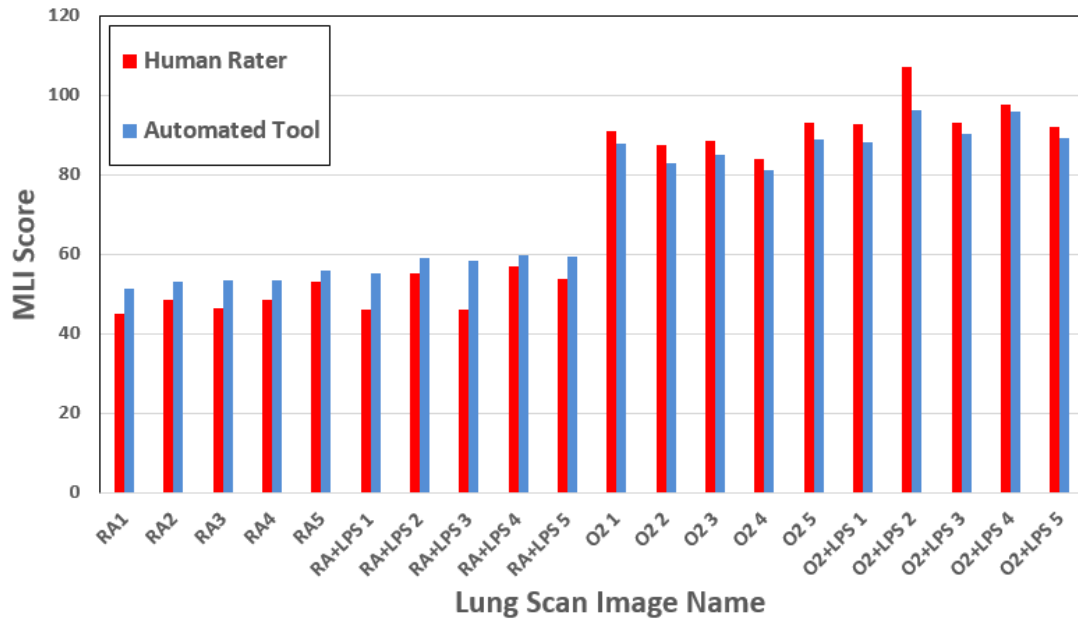


Figure 4.3 MLI scores from the proposed automated system versus the MLI score (MLI_{ind}) of the human rater.

Figure 4.4 shows a Bland-Altman plot, comparing the MLI scores from the human rater and the automated system. The mean difference in the MLI score between the two methods is 0.95 (i.e., low bias). The plot shows a negative trend between the MLI scores, with the automated system scoring lower than the human rater for images with higher MLI scores, and higher than the human rater for images with lower MLI scores.

Table 4.1 RMSD between human rater and automated system for each WSI group.

| WSI Group Name | RMSD between human rater and automated system |
|----------------------|---|
| RA | 5.35 ± 1.50 |
| RA + LPS | 7.62 ± 3.43 |
| O ₂ | 3.65 ± 0.61 |
| O ₂ + LPS | 5.61 ± 3.28 |

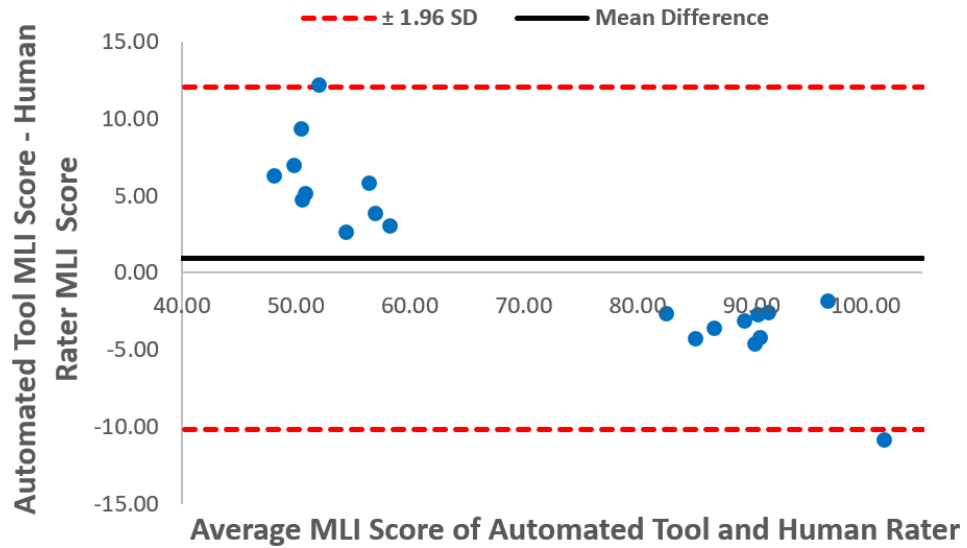


Figure 4.4 Bland-Altman plot for MLI score (MLI_{ind}) of automated system versus the human rater.

4.4 Analysis and Discussion

The MLI score of the proposed automated system was compared against the manual assessment of a human rater, showing a strong correlation $r=0.9931$. In Figure 4.4 a negative trend from the Bland-Altman analysis is noted. The WSIs which have lower MLI

scores are the RA and RA + LPS groups, which have much smaller and more frequent alveoli compared to the WSIs from the O₂ and O₂ + LPS groups.

It is possible that the human rater could have biases when determining intersections, which could have played a role in the differences between the two methods. The FOV images which were used by the automated system were different from those of the human rater, which may also explain some of the differences. Furthermore, the number of accepted FOV images used by the automated system ranged from 782 to 1846 per WSI for MLI scoring, which is much higher than the human rater (approximately 300 to 500 FOV images per WSI). A larger number of FOV images should result in a lower standard error of the MLI score, suggesting that the standard error of the automated system should be less than that of the human rater. Section 6.2 examines the relationship between the number of FOV images and standard error.

4.4.1 Mann-Whitney U Test

The MLI score is a commonly used method for quantifying lung structure and monitoring emphysematous changes to lung parenchyma. An ideal method for MLI quantification would be able to detect the differences between distinct groups, such as the four conditions in the dataset described in Chapter 4.2. Exposure to different conditions, such as high concentrations of lipopolysaccharides and oxygen, can lead to pathophysiological changes that mimic those associated with emphysema [53], [54]. While looking at the MLI scores of both the human rater and automated system from Figure 4.3, it appears that the results of the latter show better differentiation between groups compared to those of the former.

Specifically, it appears as though the automated system can better differentiate between the RA vs. RA + LPS groups compared to the human rater. Similarly, it appears that the automated system can better differentiate between the O₂ vs. O₂ + LPS groups compared to the human rater. The distribution of the MLI scores for the different groups is unknown, and therefore it is uncertain whether these scores follow a normal distribution. Therefore to test this theory, the non-parametric Mann-Whitney U test [55] is performed to determine if either the automated method or the human rater can differentiate between the groups (*i.e.*, RA vs. RA + LPS, O₂ vs. O₂ + LPS). The results of the Mann-Whitney U test are presented in Table 4.2

Table 4.2 Mann-Whitney U Test Results: Automated System vs. Human Rater ($\alpha=0.05$).

| MLI Method | Groups | Mean (MLI) | Min (MLI) | Max (MLI) | U-Calc | U-Crit | P-Value |
|------------------|----------------------|------------|-----------|-----------|--------|--------|---------|
| Automated System | RA | 53.40 | 51.30 | 55.79 | 1 | 4 | 0.016 |
| | RA + LPS | 58.30 | 55.19 | 59.85 | | | |
| Automated System | O ₂ | 85.14 | 81.20 | 88.73 | 1 | 4 | 0.016 |
| | O ₂ + LPS | 91.91 | 88.05 | 96.24 | | | |
| Human Rater | RA | 48.27 | 45.02 | 53.19 | 8 | 4 | 0.35 |
| | RA + LPS | 51.49 | 45.91 | 56.85 | | | |
| Human Rater | O ₂ | 88.74 | 83.9 | 92.95 | 3 | 4 | 0.05 |
| | O ₂ + LPS | 96.46 | 91.96 | 107.09 | | | |

Based on the results from Table 4.2, the automated system shows a significant difference between the RA (M=53.40) vs. RA + LPS (M=59.85) groups ($U(5,5) = 1$, p-value = 0.016) and a significant difference between the O₂ (M=85.14) vs. O₂ + LPS (M=91.91) groups ($U(5,5) = 1$, p-value = 0.016). The human rater, however, fails to show significant difference between the RA (M=48.27) vs. RA + LPS (M=51.49) groups ($U(5,5) = 8$, p-value = 0.35) but does show significant difference between the O₂ (M=88.74) vs. O₂ + LPS (M=96.46) groups ($U(5,5) = 3$, p-value = 0.05). The results from the Mann-Whitney test suggest that the automated system shows better differentiation between the groups compared to the human rater; however, because of the small sample size, the evidence is not strong enough to make any definitive claims.

4.4.2 Inter- and Intra-Rater Variability

MLI scoring performed by human raters is subject to inter- and intra-rater variability [37]. Since a subset of 10 WSIs (*i.e.*, 5 WSIs from RA group and 5 WSIs from O₂ + LPS group) used in this paper were the same WSIs used in [37], which involved 3 human raters, some additional analyses are possible. In [37], two of the human raters, labelled rater 1 and rater 2, are considered novice scorers, who were provided 5 hours of training by an expert scorer. Rater 3 was the same rater mentioned in 4.2.2, so their ratings are referred to as *Rater 3 Trial 1* (MLI score using the microscopic software) and *Rater 3 Trial 2* (MLI score using FOV images from [37]). In [37], human raters used FOV images that were extracted using a sliding window of size 1072×1388 pixels with 50% and 75% overlap in the horizontal and vertical directions, respectively.

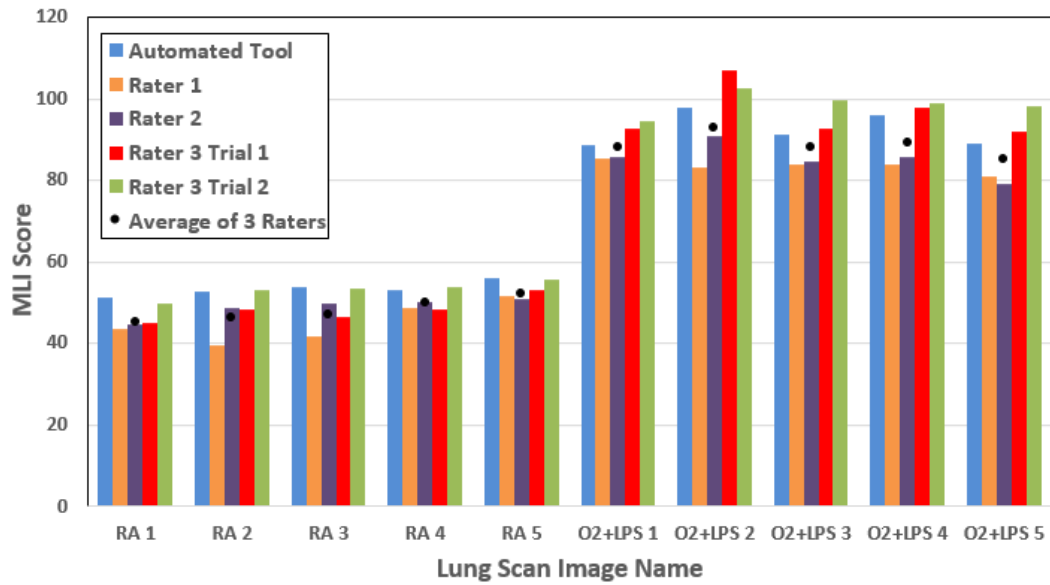


Figure 4.5 Calculated MLI scores (MLI_{ind}) from the proposed automated system and three different human raters.

Figure 4.5 shows a plot of the MLI scores from the subset of 10 WSIs, from the automated system and three human raters. For Rater 3, there is noticeable differences between their two trials, with a RMSD of 4.90 ± 1.92 . These differences show the intra-rater variability that can occur during such measurements, noting that some differences could be a result of the different FOV images used between the two scorings. When analyzing the MLI scores of the other human raters, there is noticeable variability and biases between the three human raters (Fig. 4.5; e.g., Rater 3 typically scores the WSIs higher compared to raters 1 and 2). The MLI scores from the automated system fall within this range of the human raters.

The average MLI score of the 3 raters was compared against the MLI score of the automated system. The average of the 3 raters was first computed by taking the average of

the two scores from Rater 3, and then averaging that with the scores from Raters 1 and 2. The RMSD was 4.71 ± 2.16 and there was a very strong correlation ($r=0.9955$). A Bland-Altman analysis was performed with the plot shown in Figure 4.6

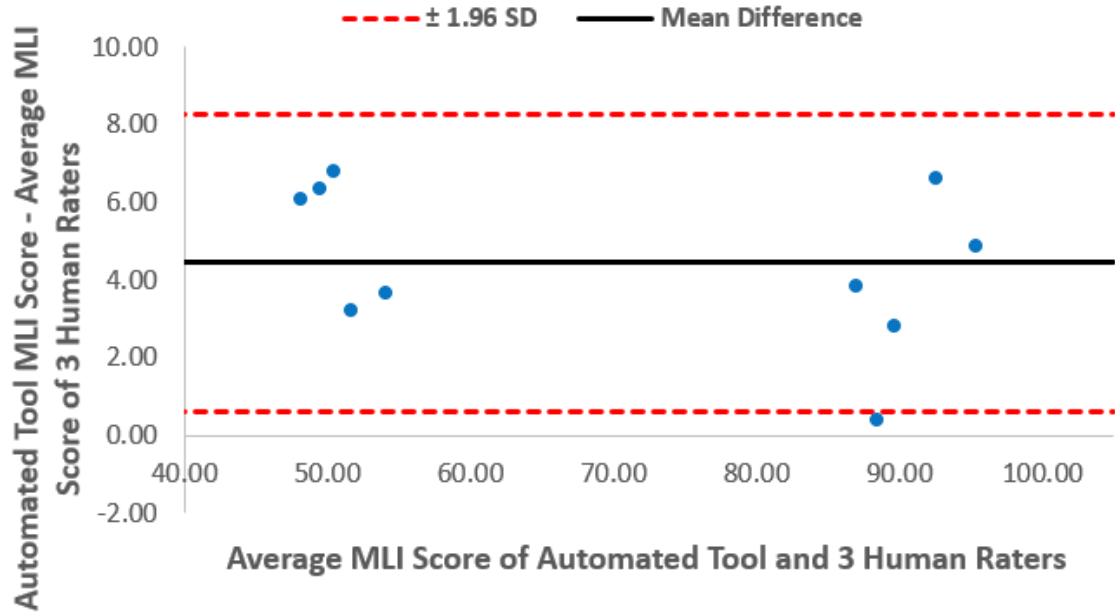


Figure 4.6 Bland-Altman plot for the MLI score (MLI_{ind}) of the automated system versus the average score of the 3 human raters.

There is a noticeably larger bias (mean difference of 4.45 in Figure 4.6 and 0.95 in Figure 4.5), while the limits of agreement are much tighter (0.61 to 8.28 in Figure 4.6 and -10.13 to 12.03 in Figure 4.5). While a negative trend was observed in Figure 4.5, there is no such trend in Figure 4.6. These results indicate that the performance of the automated system might be better than the assessment that a single rater suggests (e.g., Figure 4.5).

4.4.3 Processing Time: Automated System vs. Human Rater

The processing time required for the human rater to calculate the MLI score for a single WSI is on the order of hours [37]. The processing time required by the proposed system to calculate the MLI score for a single WSI was approximately 15 minutes. The system was tested on a standard workbook with an Intel Xeon E3-1505M v5 2.80 GHz CPU, 32 GB of installed RAM, a single NVIDIA Quadro M1000M with 4 GB memory. While minimizing processing time was not the priority in this work, the automated system is substantially faster in calculating the MLI score, with the added benefit of a fully automated pipeline. This allows a user to input multiple WSIs, which can be processed continuously unsupervised until completion.

4.5 Conclusions

An automated system to calculate the MLI score through intersection counting between the guideline and alveolar border wall was introduced. This system was evaluated using a mouse lung dataset, consisting of 20 WSIs from four different groups (i.e., RA, RA + LPS, O₂, and O₂ + LPS), comparing the automated system against the assessment of a human rater. Results were encouraging with an overall RMSD of 5.73 ± 5.65 and very strong correlation ($r=0.9931$). Results from a Mann-Whitney U test demonstrated that the MLI scores from the automated system can identify statistically significant differences between the different mouse groups (i.e., RA vs. RA + LPS and O₂ vs. O₂ + LPS), while the human rater may not. There is intra- and inter-rater variability in MLI scoring with human raters, and the proposed automated system falls within this variability. While processing time was not a priority in this work, the automated system required significantly less time (i.e., ~15 minutes) to calculate the MLI score of a WSI compared to a human rater (i.e., ~10 hours).

Chapter 5: Implementation of Direct MLI Scoring

The MLI score can be measured using the direct method. Similar to the indirect method, FOV images are extracted and a guideline is superimposed on each FOV image. FOV images are rejected if any part of the guideline intersects with the pleural space (*i.e.*, outside the lung space), a bronchus, or a blood vessel. The remaining FOV images are accepted and further analyzed, with the length of chords along the superimposed guideline measured and recorded. The MLI score is the average chord length computed across all accepted FOV images.

In this chapter, the implementation of the Direct MLI Module is presented. The pipeline of the proposed automated system using the direct MLI method is shown in Figure 5.1. This is identical to the generic pipeline shown in Figure 3.1, except for the *Chord Measurements* module, which replaces the generic *MLI Estimation* module.

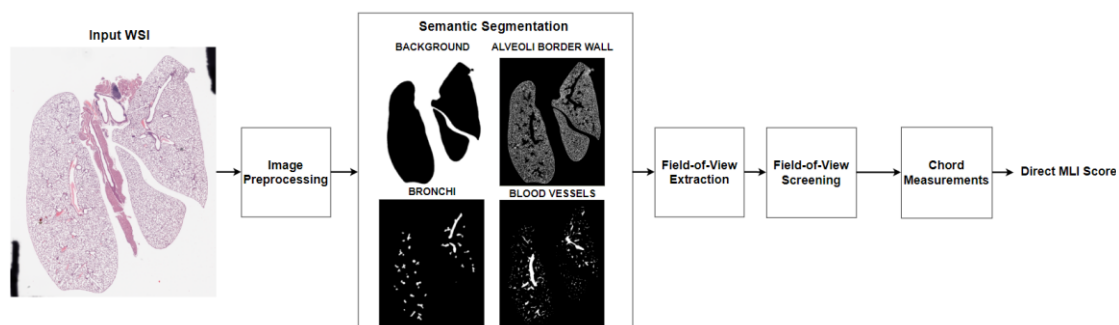


Figure 5.1 Block Diagram of the proposed automated system pipeline with direct MLI (MLI_{dir}) method implemented.

5.1 Direct MLI Module

In the direct method for MLI scoring, the lengths of chords between consecutive alveolar septa along a superimposed guideline are recorded for every accepted FOV image. The original guideline length is marked in blue, while intersections are marked in red, and chord measurements are marked in green (Figure 5.2).

If the guideline falls completely within alveoli or alveolar ducts, the guideline is extended on both sides until it reaches an alveolar septum, and the full chord length is recorded (Figure 5.2a). If either end of the superimposed guideline falls within an alveolus or alveolar duct, the guideline is extended until it reaches an alveolar septum (Figure 5.2b), and the resulting chord is measured. If the end of the guideline falls within a septum, the guideline is not extended (Figure 5.2c). If a chord extension reaches the end of a FOV image, the distance from the initial intersection to the end of the extension line is recorded, as it is better to record the measurement rather than eliminate it entirely (Figure 5.2c). Guidelines are not extended if doing so will result in contact with pleura, bronchi, or blood vessels (Figure 5.2d), and are instead essentially truncated to the nearest septum. When guidelines are not extended, that portion of the guideline remains blue in colour and does not contribute to the chord length estimation. This approach of extending the guideline provides an unbiased sampling of chords, irrespective of their length, ensuring that large alveoli are not undersampled.

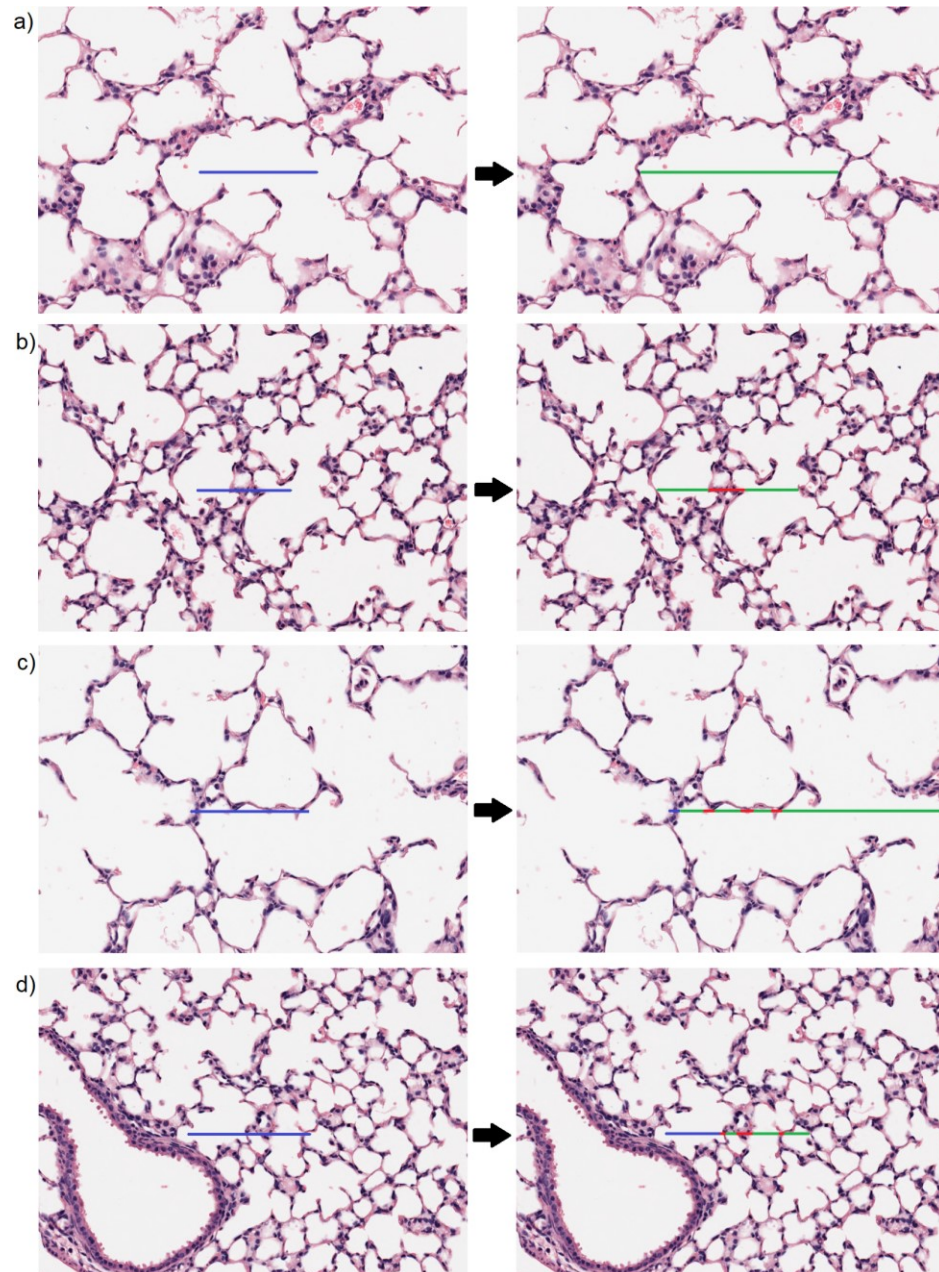


Figure 5.2 a) The guideline fits entirely inside an alveoli (left) and therefore both ends are extended until they reach an alveolar border wall (right). b) Both ends fall inside an alveoli (left) and are therefore extended until they reach an alveolar border wall (right). c) Left side of guideline falls inside an alveolar septa while the right side falls within an alveoli (left) and therefore the left side of the guideline is not extended and the right side of the guideline is extended until it reaches the edge of the FOV image, since it did not reach an alveolar septa (right). d) Left edge of the guideline falls within an alveoli (left) but is not extended because doing so would result in contact with a bronchi (right).

This chord length measurement process is repeated for every accepted FOV image, with the results saved in a CSV file (Figure 5.3). The MLI score is then calculated using Equation 2.3 from Section 2.4.2, where the sum of all chord measurements (in μm) is divided by the total number of chords measured from all accepted FOV images. The user can choose to save FOV images for verification or further analyses. FOV images can be saved with the guideline superimposed and with or without colours indicating the detected intersections and chords.

| Original Image Name | FOV Name | FOV corner x1 | FOV corner x2 | FOV corner y1 | FOV corner y2 | ACC/REJ | Image | Chord | Chord Measurements | # ACC Images | # Chords | Average Chord Length |
|---------------------|----------|---------------|---------------|---------------|---------------|---------|--------|-------|--------------------|--------------|----------|----------------------|
| O2 + LP5 3 | img_1 | 1 | 1388 | 1 | 1072 | REJ_BG | img_75 | | 53.273695 | 1758 | 4537 | 64.67643571 |
| | img_2 | 348 | 1735 | 1 | 1072 | REJ_BG | img_75 | | 112.52201 | | | |
| | img_3 | 695 | 2082 | 1 | 1072 | REJ_BG | img_75 | | 27.88156 | | | |
| | img_4 | 1042 | 2429 | 1 | 1072 | REJ_BG | img_76 | | 28.379445 | | | |
| | img_5 | 1389 | 2776 | 1 | 1072 | REJ_BG | img_76 | | 60.74197 | | | |

Figure 5.3 Example CSV file content containing all data recorded for MLI scoring using the direct method (first 5 entries). *Original Image Name* stores the name of the analyzed WSI. *FOV Name* is used to identify each extracted FOV image. The pixel locations of the two opposite corners of each FOV image are stored under the *FOV corner x1, x2, y1, y2* headings. *ACC/REJ* heading stores whether the FOV image was accepted or rejected and why. *Image Chord* stores which FOV image is being analyzed, with the chord measurement stored directly to the right under the *Chord Measurements* header. *# ACC Images* stores the total number of accepted FOV images. *# Chords* stores the total number of chords measured from all the accepted FOV images. *Average Chord Length* stores the average chord length (L_m), which is the MLI score calculated

5.2 Biases between Direct and Indirect MLI Scoring

The direct method for MLI scoring (*i.e.*, measuring chord lengths) is a more labor-intensive and time-consuming approach compared to the indirect method (*i.e.*, counting

intersections) when performed by a human rater. Measuring chords directly, however, allows the user to see the variance of the airspace measures and obtain a better understanding of alveolar surface area of the lungs. According to [38], both methods are unbiased, and when proper precautions are taken, they can provide the user with an accurate measurement of the free air space within the lung. However, there are inherent differences between the two methods, which can result in differences in MLI scores. The direct and indirect methods for MLI scoring are analyzed in this chapter using the same WSIs and extracted FOV images to investigate these differences. Specifically, two potential biases that account for the difference between the direct and indirect methods are examined: 1) *Septa Bias*, and 2) *Partial Chord Bias*.

5.2.1 Septa Bias

The indirect method is an alternative approach to the direct method to estimate the mean free distance of the air spaces within the lung (see Section 4.1). It utilizes a guideline of fixed length, and counts the number of intersections between the guideline and alveolar septa. The MLI score is calculated by summing the fixed guideline lengths from all accepted FOV images and dividing by the total number of intersections measured, as shown in Equation 2.2. This approach, however, assumes that the widths of the alveolar septa are negligible. The septa widths are essentially included as part of the chord length in the indirect method, resulting in a higher MLI score than the direct method, which only includes the length between septa. From observation alone, it is shown that the total length of the septa widths (shown in red in Figure 5.4) is not insignificant relative to the entire length of the fixed guideline.

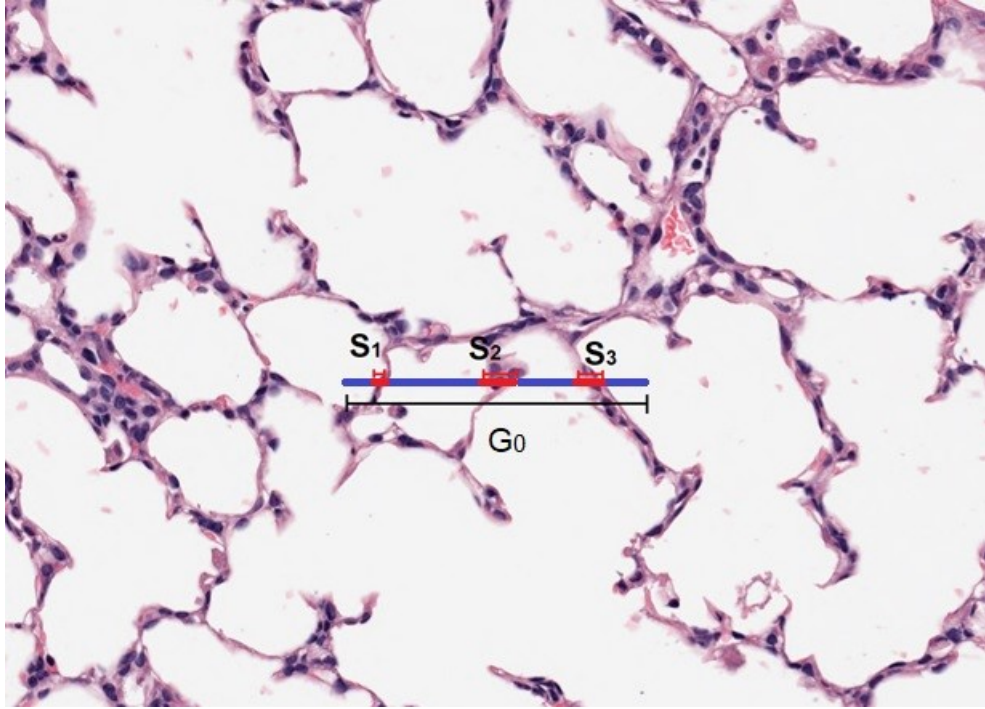


Figure 5.4 *Septa Bias* FOV image, where G_0 represents the original length of the guideline while septa widths are represented by S_1 , S_2 , and S_3 .

To remove this bias, the indirect method is modified by subtracting the septa widths of every intersection crossing from the guideline length for each FOV image, which in turn results in a lower MLI score. The MLI score with the *Septa Bias* removed is termed as MLI_{indS} (Equation 5.1) and computed as:

$$MLI_{indS} = \frac{\sum_{i=1}^N G_S(i)}{\sum_{i=1}^N C(i)} = \frac{\sum_{i=1}^N (G_0 - \sum_{j=1}^{C(i)} S_i(j))}{\sum_{n=1}^N C(i)} \quad (5.1)$$

where N is the number of FOV images that were accepted, $G_S(i)$ is the length of the guideline with all septa widths subtracted (in μm) in the i^{th} FOV image, G_0 is the original

length of the superimposed guideline (in μm), $S_i(j)$ is the width of the septa of the j^{th} intersection in the i^{th} FOV image (in μm), and $C(i)$ is the number of intersection crossings for the i^{th} FOV image.

5.2.2 Partial Chord Bias

The indirect method for MLI scoring uses a guideline of fixed length. The ends of the fixed length guideline often fall within alveolar septa, resulting in partial chords. The indirect method estimates the average chord length of a FOV image by taking the guideline length and dividing it by the number of intersections. The partial chords at the ends of the guideline result in an underestimate of the average chord length, as large alveoli are undersampled. The direct method extends the ends of the guideline to remove partial chords as described in Section 5.1. The difference caused by the partial chords is defined as the *Partial Chord Bias*.

To account for this bias, the indirect method can be modified by extending the ends of the guideline in the same manner as is done for the direct method in Section 5.1. Note that extending the guideline for the indirect MLI method is asymmetrical in that the left side of the guideline is extended until an entire septa wall is fully crossed, whereas the right side is extended until it reaches a septa wall, as shown in Figure 5.5. This asymmetrical extension allows for the number of intersections to equal the number of chords.

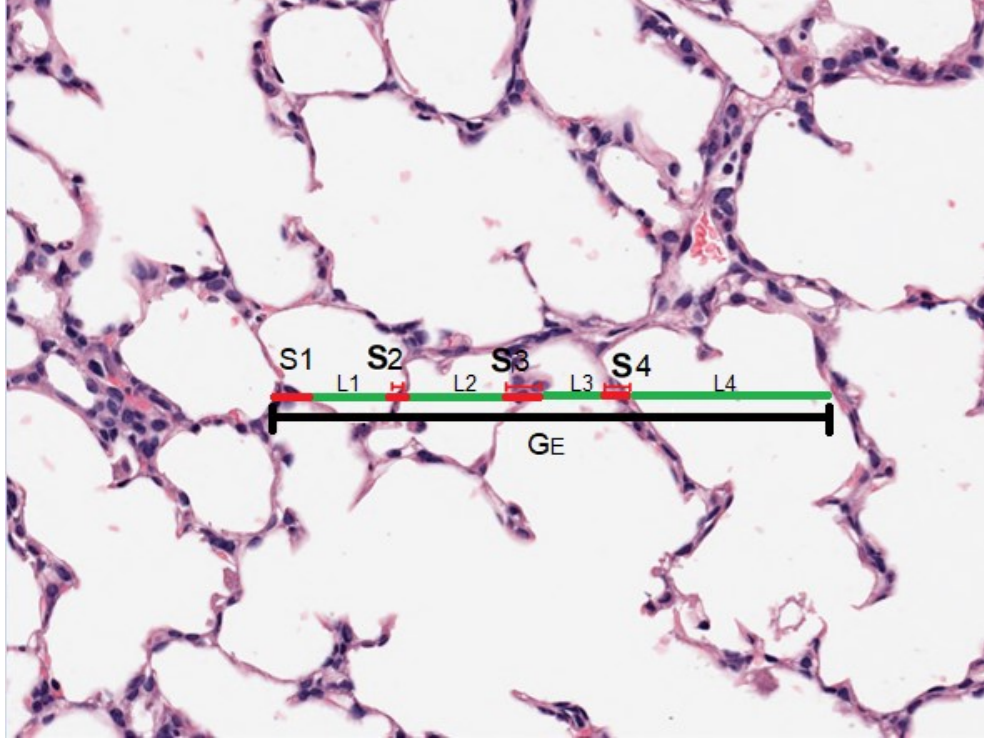


Figure 5.5 *Partial Chord Bias* FOV image, where G_E represents the total length of the new guideline, which is the sum of chord measurements (L) and septa widths (S).

The MLI score with associated *Partial Chord Bias* is now defined, which is termed as

MLI_{indE} (Equation 5.2)

$$MLI_{indE} = \frac{\sum_{i=1}^N G_E(i)}{\sum_{i=1}^N C(i)} = \frac{\sum_{i=1}^N (\sum_{j=1}^{C(i)} L_i(k) + \sum_{j=1}^{C(i)} S_i(j))}{\sum_{i=1}^N C(i)} \quad (5.2)$$

where N is the number of FOV images that were accepted, $G_E(i)$ is the length of the guideline with partial chords extended (in μm) in the i^{th} FOV image, $C(i)$ is the number of intersection crossings for the i^{th} FOV image, $L_i(k)$ is the length of the k^{th} chord in the i^{th}

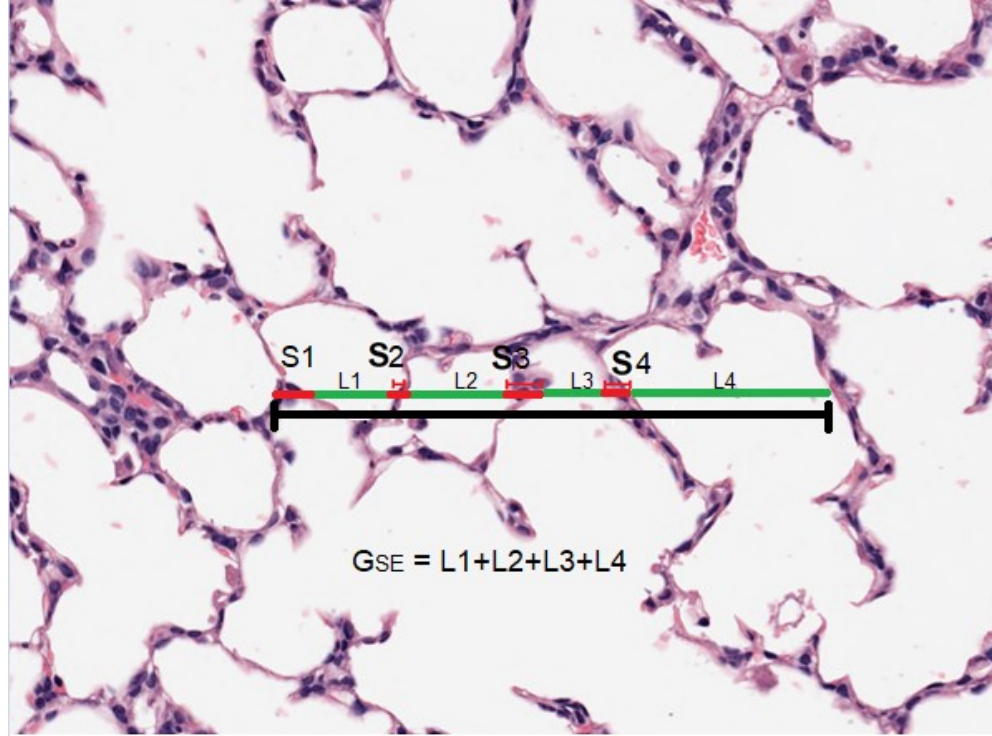


Figure 5.6 Combined *Septa Bias* + *Partial Chord Bias* FOV image, where G_{SE} represents the total length of the guideline composed of the sum of chord measurements ($L(k)$).

FOV image (in μm), and $S_i(j)$ is the width of the septa of the j^{th} intersection in the i^{th} FOV image (in μm).

5.2.3 Combined Septa Bias and Partial Chord Bias

By combining the methods to account for *Septa Bias* (MLI_{indS}) and *Partial Chord Bias* (MLI_{indE}), the MLI score from the indirect method should match the direct method. Figure 5.6 illustrates the combination of the methods to account for *Septa Bias* (MLI_{indS}) and *Partial Chord Bias* (MLI_{indE}), where there are four crossings and the guideline length used in the MLI score is the sum of the four chords shown in green, which excludes the septa widths from the guideline length.

The indirect MLI score with associated *Septa Bias* and *Partial Chord Bias* accounted for is termed MLI_{indSE} (Equation 5.3).

$$\begin{aligned}
 MLI_{indSE} &= \frac{\sum_{i=1}^N G_{SE}(i)}{\sum_{i=1}^N C(i)} = \frac{\sum_{i=1}^N [\sum_{j=1}^{C(i)} L_i(k) + \sum_{j=1}^{C(i)} S_i(j) - \sum_{j=1}^{C(i)} S_i(j)]}{\sum_{i=1}^N C(i)} \\
 &\equiv \frac{\sum_{i=1}^N \sum_{k=1}^{K(i)} L_i(k)}{\sum_{i=1}^N K(i)} = MLI_{dir} \quad (5.3)
 \end{aligned}$$

where N is the number of FOV images that were accepted, G_{SE} is the length of the guideline with both *Septa Bias* and *Partial Chord Bias* included (in μm), $C(i)$ is the number of intersection crossings for the i^{th} FOV image, $S_i(j)$ is the width of the septa of the j^{th} intersection in the i^{th} FOV image (in μm), $L_i(k)$ is the length of the k^{th} chord in the i^{th} FOV image (in μm), and $K(i)$ is the number of chords for the i^{th} FOV image.

The number of intersections crossings, $C(i)$, is equivalent to the number of chords, $K(i)$, as a result of the asymmetrical extension for each FOV image. The indirect MLI score with *Septa Bias* and *Partial Chord Bias* (MLI_{indSE} ; Equation 5.3) is therefore shown to be equivalent to the direct method for MLI scoring (MLI_{dir} ; Equation 2.3)

5.3 Methodology

5.3.1 Histopathology Dataset of Lung Tissue

To evaluate the direct MLI module and investigate the effects of the *Septa Bias* and *Partial Chord Bias*, the same 20 histopathology WSIs of mouse lung tissue as described in Section

4.2.1 is used, consisting of four distinct groups (RA group, RA + LPS group, O₂ group, and O₂ + LPS group).

5.3.2 MLI Scoring: Indirect Method versus Direct Method

The proposed automated system described in chapter 3, using the indirect method (MLI_{ind}) from Section 4.1, and the direct MLI module (MLI_{dir}) from Section 5.1, are used to calculate the MLI score for all 20 WSIs described in Section 5.3.1. A modified version of the indirect MLI module from Section 4.1, and the automated system described in chapter 3, are used to assess the impact of *Septa Bias* on the resulting MLI score. Through the modification, the widths of the intersection crossings (in μm) identified in each FOV image are measured. By subtracting the septa widths from the original guideline length in each accepted FOV image, the MLI score (MLI_{inds}) can be calculated as per equation 5.1 (illustrated in Figure 5.4). To assess the impact of *Partial Chord Bias*, the direct module from Section 5.1, and the fully automated system described in Chapter 3 is used. Through the modification, the resulting guideline length (in μm) for each FOV image was the sum of chord lengths and septa widths measured (illustrated in Figure 5.5). The resulting MLI score (MLI_{indE}) can then be calculated as per equation 5.2. The MLI score (MLI_{indSE}) is also computed, which takes into account both the *Septa Bias* and *Partial Chord Bias*.

FOV images of size 1072 (height) \times 1388 (width) pixels were extracted using a sliding window, with a stride length of 268 and 347 in the horizontal and vertical directions, respectively (75% overlap). Therefore, the same FOV images used in Section 4.2.2 to calculate the MLI score for the indirect method, are used to evaluate the direct MLI score

and assess the impact of *Septa Bias* and *Partial Chord Bias*. The horizontal guideline length is maintained at 312 pixels, corresponding to a length of 155.34 μm for the given scanning resolution.

5.3.3 Analysis

The MLI score of the automated system using the direct method was compared against the MLI score of the automated system using the indirect method from Section 4.3. A root-mean-squared deviation (RMSD), a Bland Altman analysis [50], and a Pearson correlation coefficient is used in the analysis.

5.4 Results

5.4.1 Direct MLI Scoring

While both indirect and direct methods can be used to calculate the MLI score, the direct approach allows the user to obtain the variance of the chord measurements and therefore a better understanding of the alveolar surface area of the lungs. Figure 5.7 displays the distribution of the chord measurements taken from one of the WSIs of the dataset.

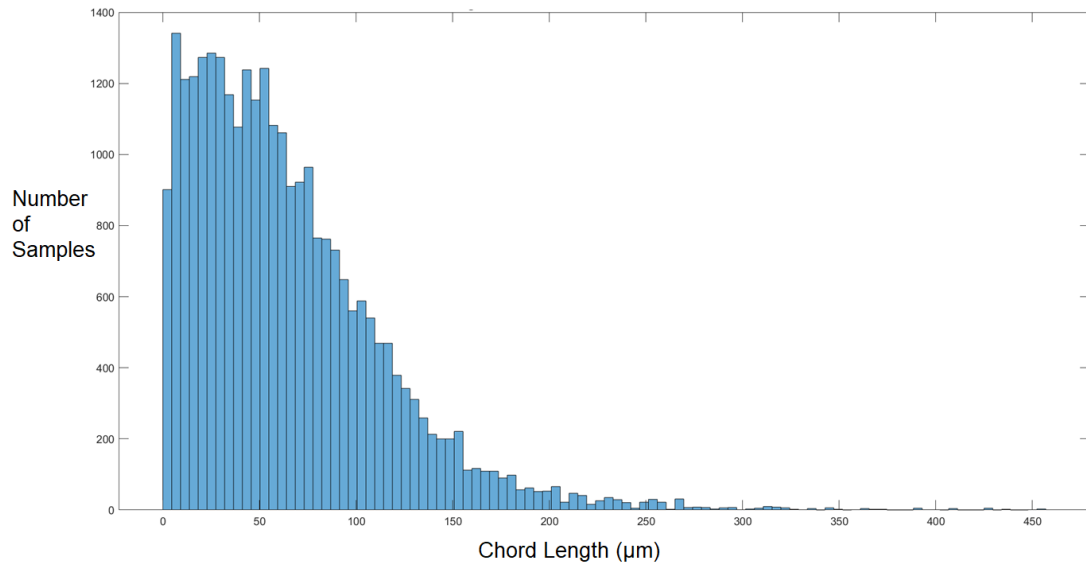


Figure 5.7 Example chord length distribution obtained from automated direct MLI (MLI_{dir}) scoring.

5.4.2 Indirect versus Direct Method for MLI scoring

The resulting MLI scores from the automated system using the indirect and direct methods are shown in Figure 5.8. The MLI scores from the direct method (MLI_{dir}) are consistently lower than the MLI scores from the indirect method (MLI_{ind}). Table 5.1 summarizes the RMSD between the MLI scores of the direct and indirect method. There was a very strong correlation ($r=0.9919$) between the MLI scores of the indirect and direct method.

Table 5.1 RMSD between the indirect and direct MLI scores.

| WSI Group Name | RMSD between human rater and automated system |
|----------------------|---|
| RA | 24.10 ± 1.17 |
| RA + LPS | 25.75 ± 1.39 |
| O ₂ | 28.38 ± 1.64 |
| O ₂ + LPS | 29.78 ± 3.13 |

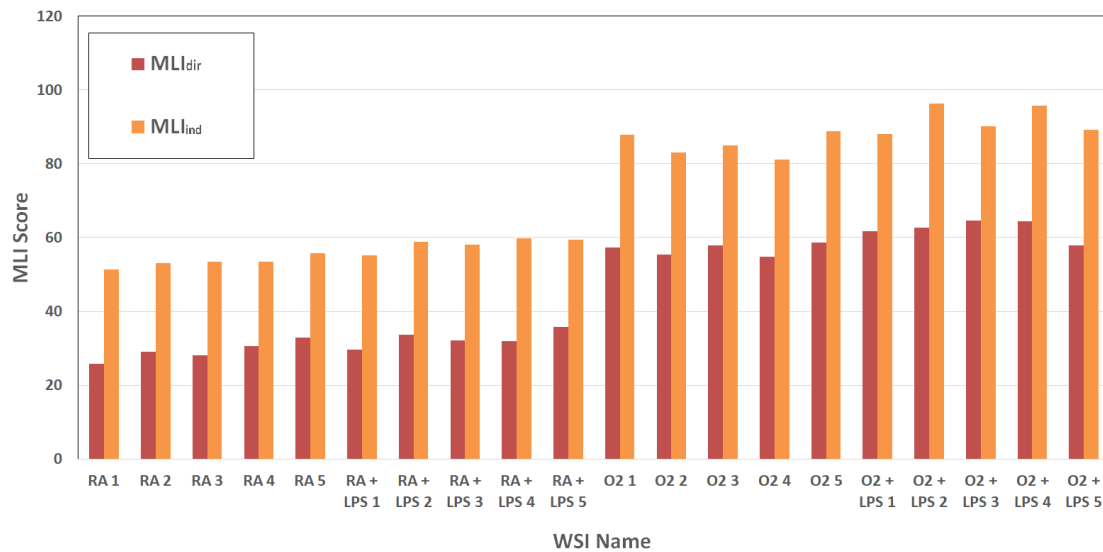


Figure 5.8 MLI score calculated by the automated system using the indirect (MLIind) and direct (MLIdir) MLI modules.

Figure 5.9 shows a Bland-Altman plot, comparing the MLI scores of the indirect and direct method. The mean difference in the MLI score between the two methods is 26.93 (i.e., high

bias). The plot shows a positive trend between the MLI scores, meaning that the difference between the indirect and direct methods is higher for images with higher MLI scores, and lower for images with lower MLI scores.

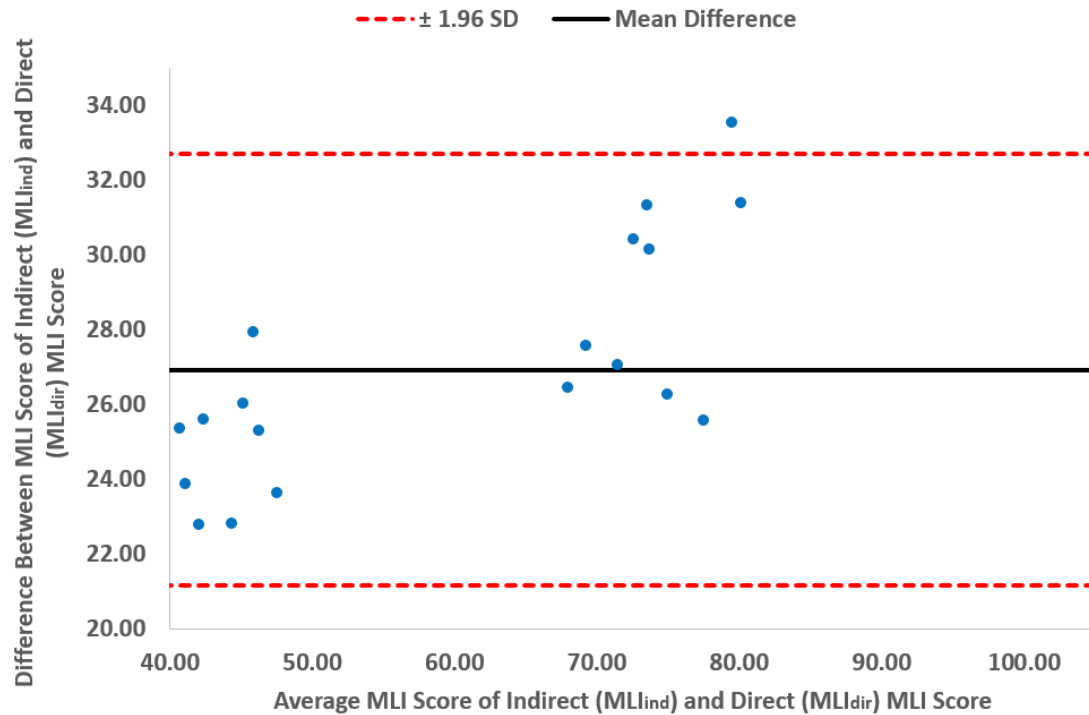


Figure 5.9 Bland-Altman plot for MLI score for indirect (MLI_{ind}) and direct (MLI_{dir}) methods.

5.4.3 Septa Bias, Partial Chord Bias, and Combined Septa Bias + Partial Chord Bias

The resulting MLI scores considering the *Septa Bias*, *Partial Chord Bias*, and combined *Septa Bias + Partial Chord Bias* described in sections 5.2.1, 5.2.2, and 5.2.3, respectively, are shown in Table 5.2, with the results graphed in Figure 5.10. When the

Septa Bias or *Partial Chord Bias* are accounted for, the MLI scores for the indirect method decrease towards the MLI score of the direct method. When both the *Septa Bias* and *Partial Chord Bias* are accounted for simultaneously, the MLI scores are equal between the indirect and direct methods.

Table 5.2 Average MLI Score per Group for Direct (MLIdir), Indirect (MLIind), *Septa Bias* (MLIindS), *Partial Chord Bias* (MLIindE), and *Septa Bias + Partial Chord Bias* (MLIindSE) MLI Scores

| WSI Group | MLIdir | MLIind | MLIindS | MLIindE | MLIindSE |
|----------------------|--------|--------|---------|---------|----------|
| RA | 29.33 | 53.40 | 35.63 | 43.22 | 29.33 |
| RA + LPS | 32.59 | 58.30 | 39.91 | 47.06 | 32.59 |
| O ₂ | 56.80 | 85.14 | 66.69 | 71.94 | 56.80 |
| O ₂ + LPS | 61.51 | 89.92 | 70.34 | 78.34 | 61.51 |

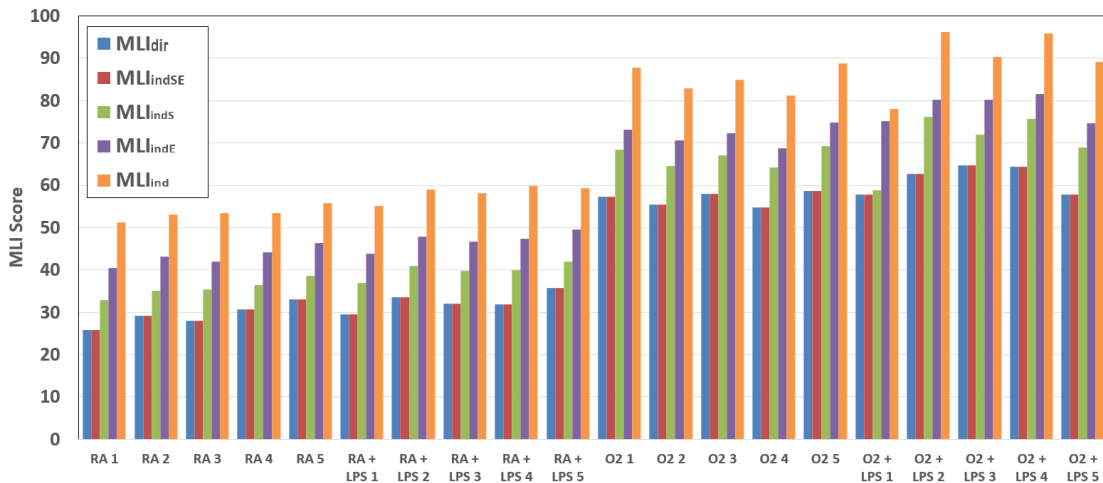


Figure 5.10 MLI Score using indirect (MLIind), direct (MLIdir), *Septa Bias* (MLIindS), *Partial Chord Bias* (MLIindE), and *Septa Bias + Partial Chord Bias* (MLIindSE) equations.

5.5 Analyses and Discussion

From Figure 5.8, we see a large difference in the MLI scores between the direct and indirect method, but still a high correlation ($r=0.9858$). Although there is a high RMSD of 27.09, there is a small standard deviation of 0.71. This suggests that although the indirect and direct methods differ greatly in MLI score, this difference is due to a bias between the two measurements, rather than an error in either of the two methods. These biases have been defined as *Septa Bias* and *Partial Chord Bias*.

By removing the widths of the septa from the length of the guideline (*Septa Bias MLI*; Eq. 5.1), there is a large drop in MLI score (*i.e.*, average drop of 18.54 μm between the indirect method and the *Septa Bias*), bringing the indirect MLI score closer to the direct MLI score. This demonstrates that the widths of the septa are not negligible and significantly impact the resulting calculation. Similarly, by extending the edges of the guideline (*Partial Chord Bias MLI*; Eq. 5.2) until an alveolar septa is reached on one end and an extra intersection is measured on the other, there is a decrease in the MLI score when compared to the indirect method (*i.e.*, average drop of 11.54 μm between the indirect method and the *Partial Chord Bias*). Finally, when both the *Septa Bias* and *Partial Chord Bias* are combined (*Septa Bias + Partial Chord Bias MLI*; Eq. 5.3), the resulting MLI score is identical to that of the direct method. The differences between the indirect and direct MLI methods are therefore a result of the *Septa Bias* and *Partial Chord Bias*, as predicted in Equation 5.3.

Table 5.3 Mann-Whitney U Test Results: Direct (MLI dir) Method for MLI Score ($\alpha=0.05$).

| MLI Method | WSI Groups | Mean (MLI) | Min (MLI) | Max (MLI) | U-Calc | U-Crit | P-Value |
|--|----------------------|------------|-----------|-----------|--------|--------|---------|
| Direct (MLIdir) Method Automated System | RA | 29.32 | 25.83 | 32.97 | 4 | 4 | 0.08 |
| | RA + LPS | 32.45 | 29.47 | 35.41 | | | |
| Direct (MLIdir) Method Automated System | O ₂ | 56.08 | 54.76 | 57.50 | 2 | 4 | 0.03 |
| | O ₂ + LPS | 61.42 | 57.25 | 63.62 | | | |

A Mann-Whitney U test is used to determine if the implementation of the direct MLI method can also differentiate between the different groups of the dataset. The results of the Mann-Whitney U test are presented in Table 5.3

Based on the results from Table 5.3, the direct MLI method (MLIdir) does not show a significant difference between the RA (M=29.32) vs. RA + LPS (M=32.45) groups ($U(5,5) = 4$, p-value = 0.08) but does show a significant difference between the O₂ (M=56.08) vs. O₂ + LPS (M=61.42) groups ($U(5,5) = 2$, p-value = 0.03). This is in contrast with the results of the Mann-Whitney U test for the indirect MLI method (MLIind) from Table 4.2, which showed significant differences between all WSI groups. These results could be a result of the limited dataset used for the Mann-Whitney U test and therefore because of the limited sample size, the evidence is not strong enough to make any definitive claims.

5.6 Conclusion

The MLI score can be measured directly through measurements of chord lengths between consecutive alveolar septa. The automated system was evaluated utilizing the *Chord Measurements* module on 20 WSIs, consisting of 4 different groups. The results of the Mann-Whitney U test demonstrate that the MLI scores from the direct method can identify statistically significant differences between the different mouse groups of the dataset. The direct method shows high correlation ($r=0.9858$) with the results from the indirect method (Chapter 4.4). The RMSD between the direct method and indirect method was large ($\text{RMSD} = 26.96$); however, there was a small standard deviation (0.71), suggesting that there is a large bias causing the difference between the two methods.

In this chapter, two biases were defined, the *Septa Bias* (Equation 5.1) and *Partial Chord Bias* (Equation 5.2), which are considered as the main contributions to the difference between the direct (MLIdir) and indirect (MLIind) MLI scores. When these biases are accounted for the MLI scores are identical between the indirect and direct methods.

In semi-automated approaches, the direct method for MLI scoring is more time-consuming and laborious compared to indirect MLI scoring. In the automated approach, however, both indirect and direct MLI scoring methods can be done in a less amount of time. When proper precautions are taken, both methods can provide an assessment of lung air space size. The direct method, however, may provide a richer amount of information that could be useful clinically. For example, the MLI score from the direct method provides the user with the

variance of air space size (*e.g.*, chord length distribution), while the indirect method only provides a single score.

Chapter 6: Effect of Number of Accepted FOV Images on MLI Score

The number of accepted FOV images is a parameter that affects the MLI score. In previous works, it is recommended that at least 300 FOV images should be used for accurate MLI scoring [37], [38]. Examining the effect of this parameter would be tedious and time-consuming with semi-automated MLI scoring techniques; however, with automated MLI scoring, such investigation is made easier. To the best of the author's knowledge, there have not been any studies systematically researching the impact of number of accepted FOV images on a dataset of real WSIs of lung tissue. Similarly, there is no work that examines the impact of number of accepted FOV images on the resulting MLI score for both indirect and direct methods simultaneously.

Crowley *et al.* examined the effects of number of chords on the resulting MLI score, and estimated that approximately 174-791 chords were required [36], depending on the actual MLI score. They did not, however, perform tests on WSIs of lung tissue, but instead on artificially created circles. These circles may not accurately replicate the complexity and variability of biological structures found in lung tissue; therefore, caution must be taken when interpreting their results. The objective of this chapter is to quantify the relationship that number of accepted FOV images has on the resulting MLI score for both indirect and direct MLI methods.

6.1 Number of Accepted FOV Images

In chapters 4 and 5, the number of accepted FOV images varied from 782 to 1846 images. With a small number of FOV images, it is expected that the calculated MLI score would

vary dramatically and therefore have a larger standard error. As the number of accepted FOV images increases, it is expected that the standard error decrease for both the direct and indirect methods for MLI calculation. Since the direct method for MLI scoring measures chord lengths directly, it is expected to see a smaller standard error for the same number of accepted FOV images compared to the indirect method. With the fully automated system, such analysis can easily be implemented.

6.2 Methodology

6.2.1 Histopathology Dataset of Lung Tissue

To quantify the impact of number of accepted FOV images has on the MLI score, the same 20 histopathology WSIs of mouse lung tissue as described in Section 4.2.1 is used, which consists of four distinct mouse groups (RA group, RA + LPS group, O₂ group, and O₂ + LPS group).

6.2.2 MLI Scoring

The proposed automated system described in chapter 3 and the *Intersection Crossing* module and *Chord Measurements* modules from Section 4.1 and 5.1, respectively, are used to calculate the MLI score for all 20 WSIs described in Section 6.2.1.1. FOV images of size 1072 (height) \times 1388 (width) pixels were extracted with a stride length of 107 and 139 in the horizontal and vertical directions, respectively (90% overlap). The guideline length is set to 312 pixels (155.34 μ m) for both the direct and indirect methods. The number of accepted FOV images varied from 2168 to 5320. For each WSI in the dataset, the MLI score was calculated through both indirect and direct methods using varying numbers of

accepted FOV images (*i.e.*, 10, 50, 100, 200, 300, 400, and 1000). This was repeated 1000 times for all WSIs at each number of accepted FOV images, allowing for the standard error to be calculated from the distribution.

6.3 Results

MLI scores are calculated for each WSI at varying numbers of accepted FOV images. This process is repeated 1000 times and the standard error for each number of accepted FOV images is calculated from the distribution. The standard errors for the WSIs in the same group are averaged, which is repeated for every number of accepted FOV images, with the results shown in Table A1 of Appendix A. The average mean and standard error for the RA, RA + LPS, O₂, and O₂ + LPS groups are plotted in Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4, respectively.

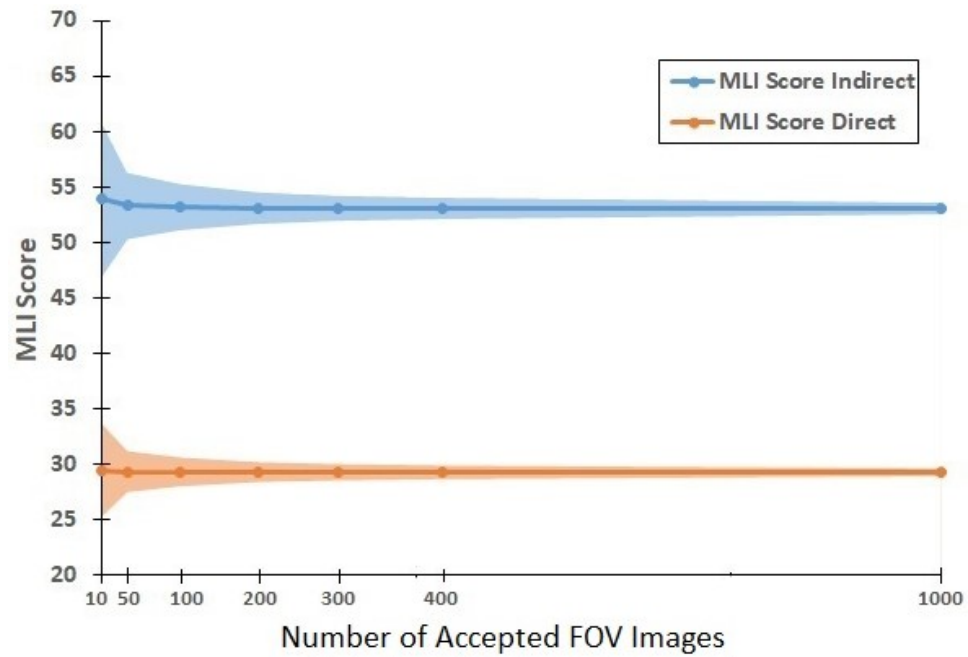


Figure 6.1 Average mean MLI score and standard error for varying number of accepted FOV images for RA group

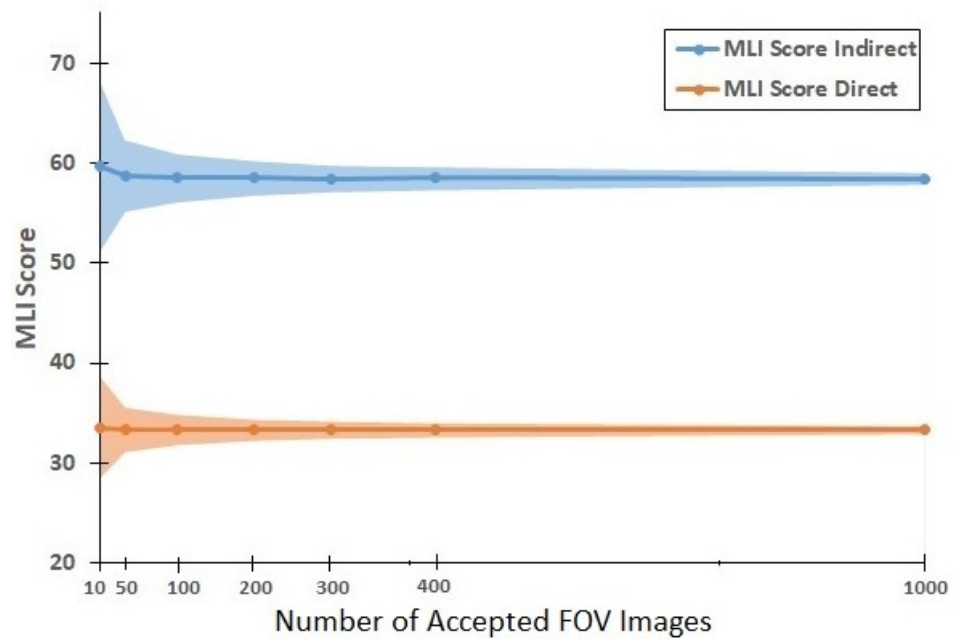


Figure 6.2 Average mean MLI score and standard error for varying number of accepted FOV images for RA + LPS group

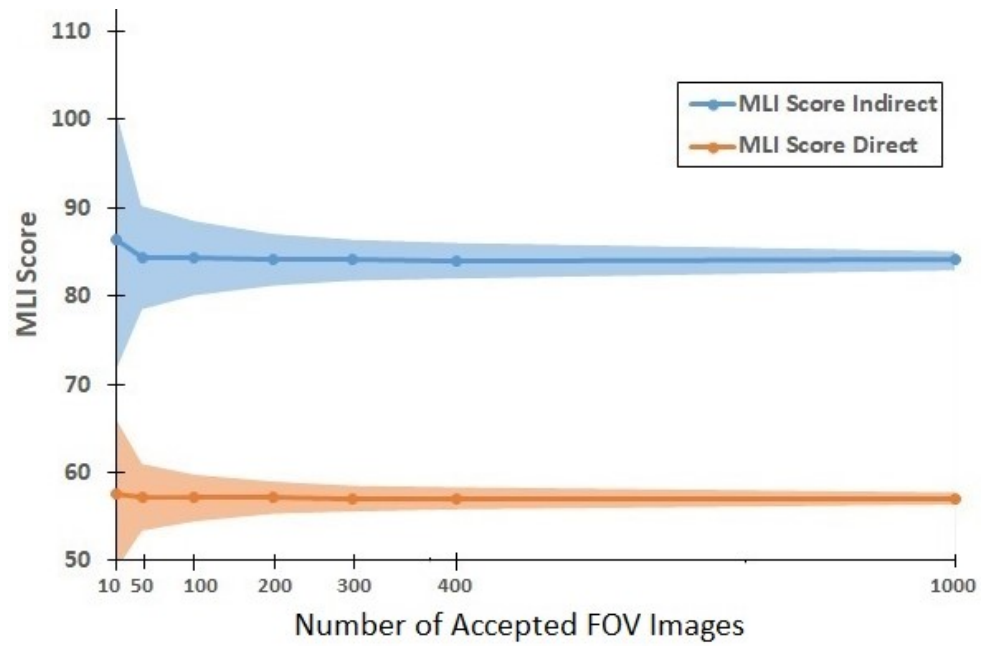


Figure 6.3 Average mean MLI score and standard error for varying number of accepted FOV images for O₂ group

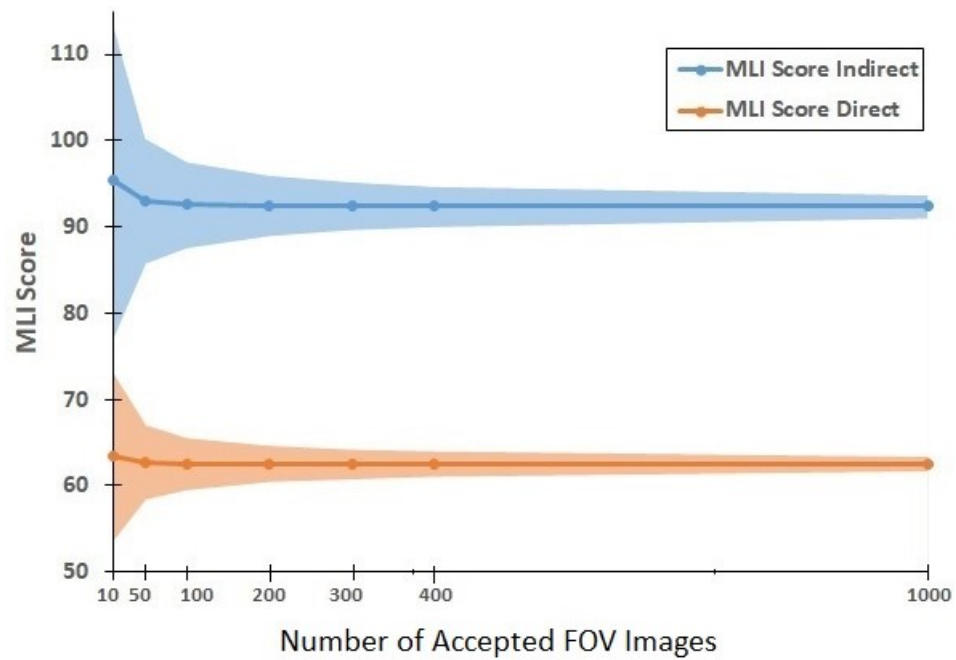


Figure 6.4 Average mean MLI score and standard error for varying number of accepted FOV images for O₂ + LPS group

6.4 Analyses and Discussion

As shown in Figure 6.1 to Figure 6.4, as the number of accepted FOV images used to calculate the MLI score increases, the standard error decreases. There is a large variability in the associated MLI score for both direct and indirect measurements when using a number of accepted FOV of 200 (*i.e.*, standard error as high as 3.54 and 2.10 for direct and indirect methods, respectively). The average standard error when using 300 accepted FOV images ranged from 0.73 to 1.72 and 1.17 to 2.81 for the direct and indirect methods, respectively, depending on the WSI group. Comparing the intra-rater variability measured in Chapter 4.5, there is a RMSD of 4.90 ± 1.92 , meaning the standard error using 300 accepted FOV images appears to be much more consistent than individual human rater analysis. The average standard error when using 1000 accepted FOV images ranged from 0.34 to 0.83 and 0.57 to 1.39 for the direct and indirect methods, respectively, depending on the WSI group. In this work, the inter- and intra-rater variability that was found during human rater analysis from [37] was compared against the standard error associated with varying the number of accepted FOV images. In future work, determining that threshold of acceptability regarding the standard error could help determine minimum requirements for the number of accepted FOV images when calculating the MLI score. The standard error of the indirect method is higher compared to the direct method for every number of accepted FOV images tested. This further demonstrates the benefit of utilizing the direct method for calculating the MLI score compared to the indirect method. The direct method results in a lower standard error at the same number of accepted FOV images, while also providing the user with the variance of chord measurements taken throughout the WSI.

6.5 Conclusion

The fully automated system was used to assess how number of accepted FOV images impacts the resulting MLI score for both indirect and direct methods. For a small number of accepted FOV images (*i.e.*, approximately below 200), there is a high standard error, shown through repeated random sampling. As the number of accepted FOV images used to calculate the MLI score increases, there is a continuous decrease in the standard error of repeated measurements for both indirect and direct MLI scoring methods. The direct method has a smaller standard error when compared to the indirect method, perhaps due to the nature of measuring chords directly, rather than estimation through intersection crossings. These results suggest that 300 accepted FOV images, as suggested in the relevant literature (*i.e.*, [37], [38]), is an ample number to ensure a low standard error in the resulting MLI score. As the number of accepted FOV images increases past 300, the standard error continuous to decrease, albeit at a slow rate. With the fully automated system, a user can easily increase the number of accepted FOV images to reduce the resulting standard error for both indirect and direct methods without the large increase to analysis time that would be found in semi-automated approaches. The results also suggest that the direct method for MLI scoring is a much more preferable approach when compared to the indirect method, due to its lower standard error for the same number of accepted FOV images.

Chapter 7: Effect of Guideline Length on MLI Score

In the indirect and direct methods for MLI scoring, a superimposed guideline is used to either count intersection crossings or record chord measurements, respectively. Guideline length is a parameter that affects the MLI score. Examining the effect of this parameter would be tedious and time-consuming with semi-automated MLI scoring; however, with automated MLI scoring, such investigation is made easier. To the best of the author's knowledge, there have not been any studies systematically researching the impact of guideline length on the resulting MLI score in histopathology images of lung tissue. Until this point, the guideline length was fixed at a length of 155.34 μm when calculating the MLI score with the automated system. In this chapter, the relationship that guideline length has on the resulting MLI score for both indirect and direct MLI methods is quantified.

7.1 Guideline Length

Until now, the length of the guideline used for the indirect and direct automated MLI scoring was fixed at 155.34 μm . In the indirect method, the guideline stays at this fixed length, and the number of intersections between the guideline and alveolar septa is counted. Increasing the guideline length would result in more intersections per FOV image (Figure 7.1a), while decreasing the guideline length would result in less intersections per FOV image (Figure 7.1b). However if a guideline length is too small, the guideline would have a lower probability of having intersections, particularly for larger alveoli, resulting in a higher and inaccurate MLI score. For indirect MLI measurements, it is likely that the length of the guideline can impact the resulting MLI scores, especially for much smaller lengths.

For the direct method, although the initial guideline length is set by the user, edges of the guideline are often extended (if they fall within alveoli or alveolar ducts) to ensure that full chords are measured. Such an approach would therefore limit the effects that guideline length has on the resulting MLI score. Increasing the initial guideline length would result in more chords measured per FOV image (Figure 7.1c), which should only increase the number of chords measured per WSI and provide a better measure of the MLI score. Decreasing the guideline length would result in less chords measured per FOV image (Figure 7.1d), but with sufficient number of FOV images, the resulting impact on the final MLI score should be small.

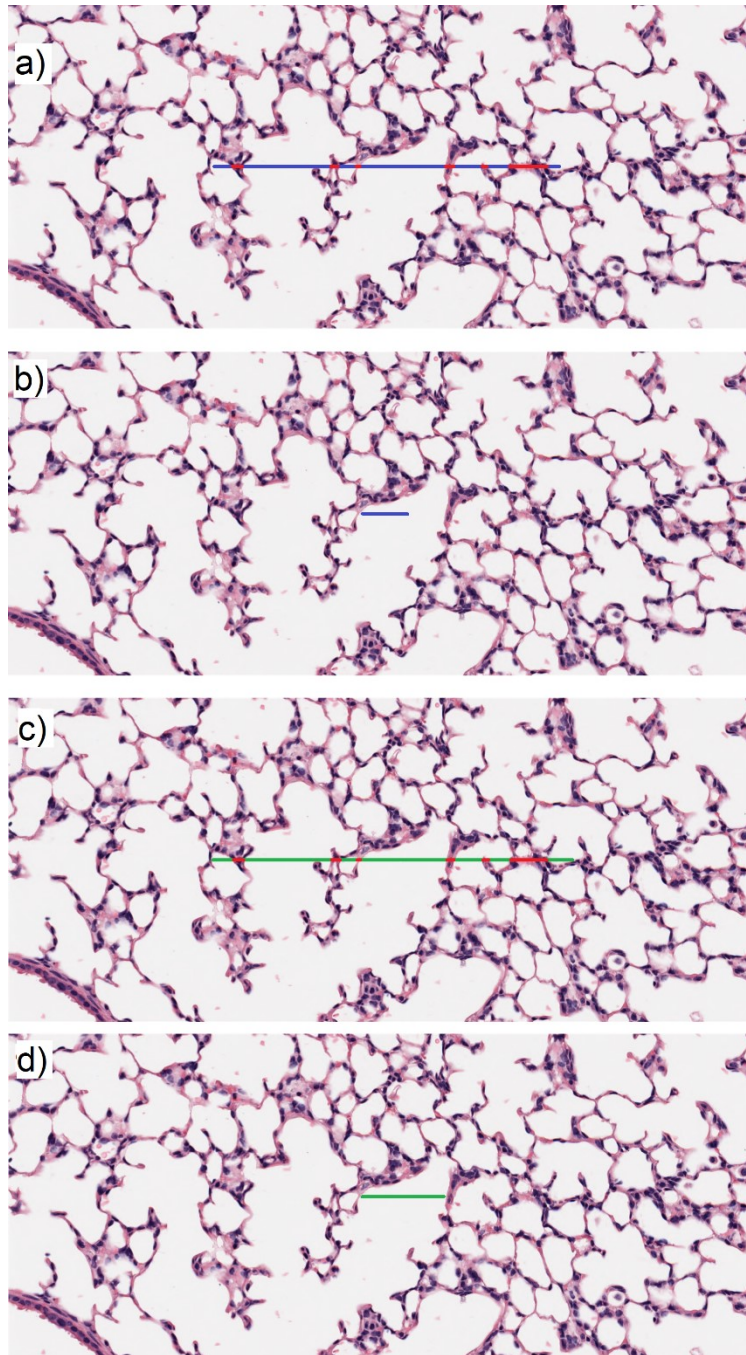


Figure 7.1 Example FOV image for a) indirect MLI scoring with large guideline, b) indirect MLI scoring with small guideline, c) direct MLI scoring with large guideline, d) direct MLI scoring with small guideline.

7.2 Methodology

7.2.1 Histopathology Dataset of Lung Tissue

To quantify the impact of guideline length on resulting MLI score, the same 20 histopathology WSIs of mouse lung tissue as described in Section 4.2.1 are used, which consists of four distinct groups (RA group, RA + LPS group, O₂ group, and O₂ + LPS group).

7.2.2 MLI Scoring

The proposed automated system described in chapter 3 and the *Intersection Crossing* module and *Chord Measurements* modules from Section 4.1 and 5.1, respectively, are used to calculate the MLI score for all 20 WSIs described in Section 4.2.1. FOV images of size 1072 (height) \times 1388 (width) pixels were extracted using a sliding window, with a stride length of 268 and 347 in the horizontal and vertical directions, respectively (75% overlap). The horizontal guideline length is varied from 39 to 702 pixels, corresponding to a length of 19.42 to 349.52 μ m, respectively, for the given scanning resolution. The MLI score is calculated for each FOV image in the dataset, for both the indirect and direct methods.

To estimate the standard error associated with different guideline lengths, FOV images of size 1072 (height) \times 1388 (width) pixels are extracted using a sliding window, with a stride length of 107 and 139 in the horizontal and vertical directions, respectively (90% overlap). For each WSI in the RA and O₂ + LPS groups of the dataset, the MLI score was calculated through both indirect and direct methods by randomly sampling 400 accepted FOV images

at each guideline length. This was repeated 1000 times for each WSI so that the standard error could be calculated from the distribution.

7.3 Results

Figure 7.2 shows the graph of the indirect MLI score and number of intersections versus varying guideline lengths.

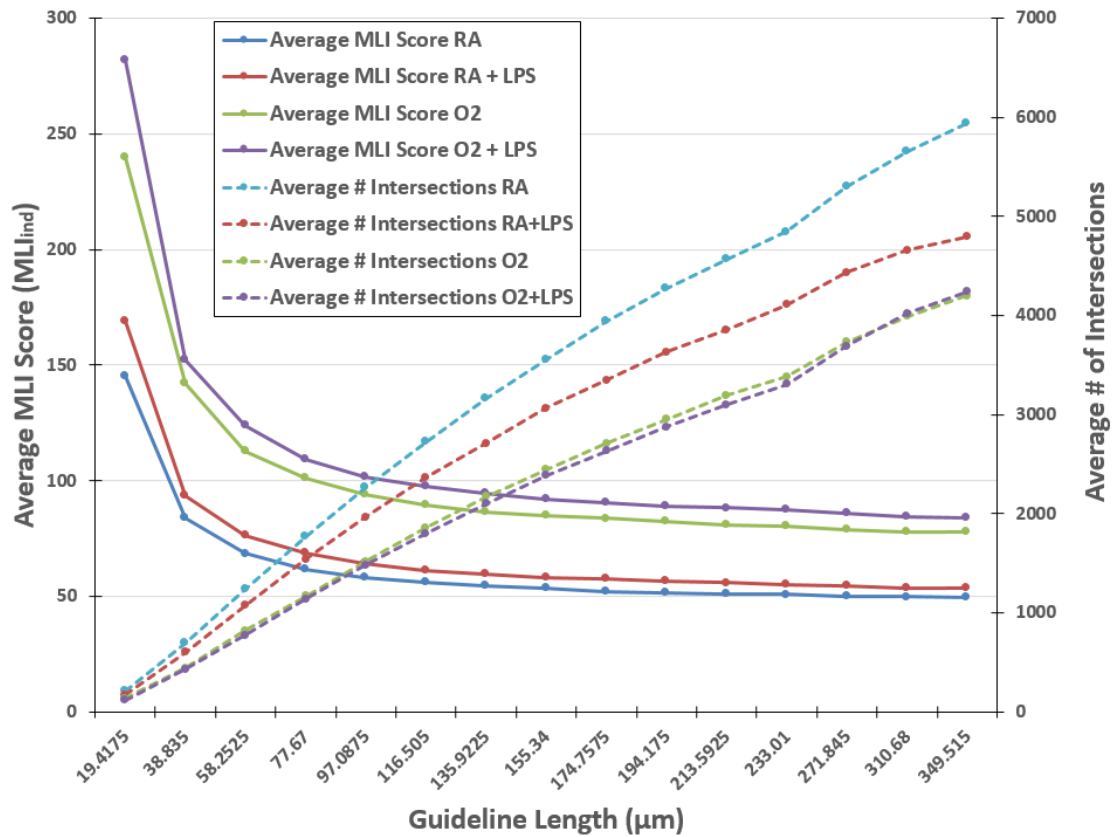


Figure 7.2 Average indirect MLI (MLI_{ind}) score (solid line) and average number of intersections (dashed line) vs. guideline length (in μm).

Figure 7.3 shows the graph of the direct MLI score and number of chords versus varying guideline lengths.

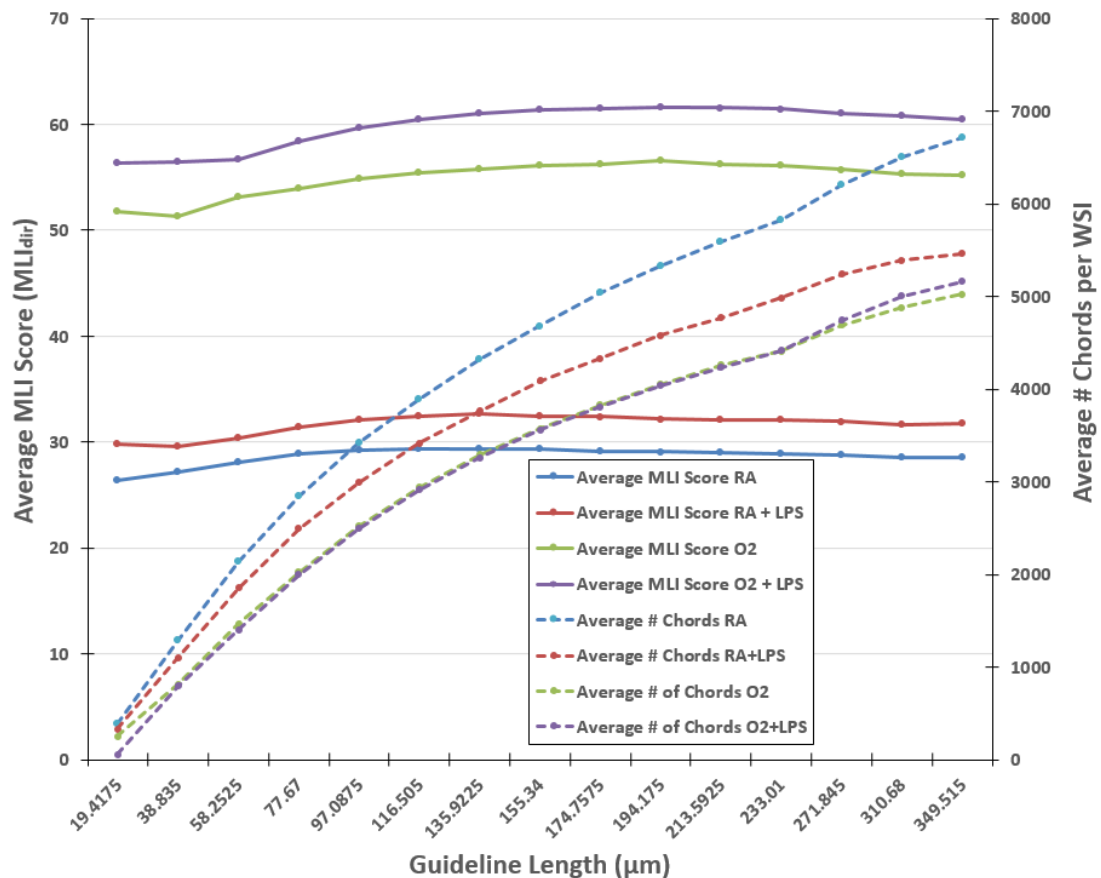


Figure 7.3 Average Direct MLI (MLI_{dir}) Score (solid line) and Average Number of Chords (dashed line) vs. Guideline Length (in μm).

As the length of the guideline increases, so does the likelihood of rejecting a FOV image due to a higher likelihood of contact between the guideline and a bronchi, blood vessel, or pleural space. Therefore while the number of extracted FOV images remains the same for each WSI, it is expected to see fewer accepted FOV images (Figure 7.4).

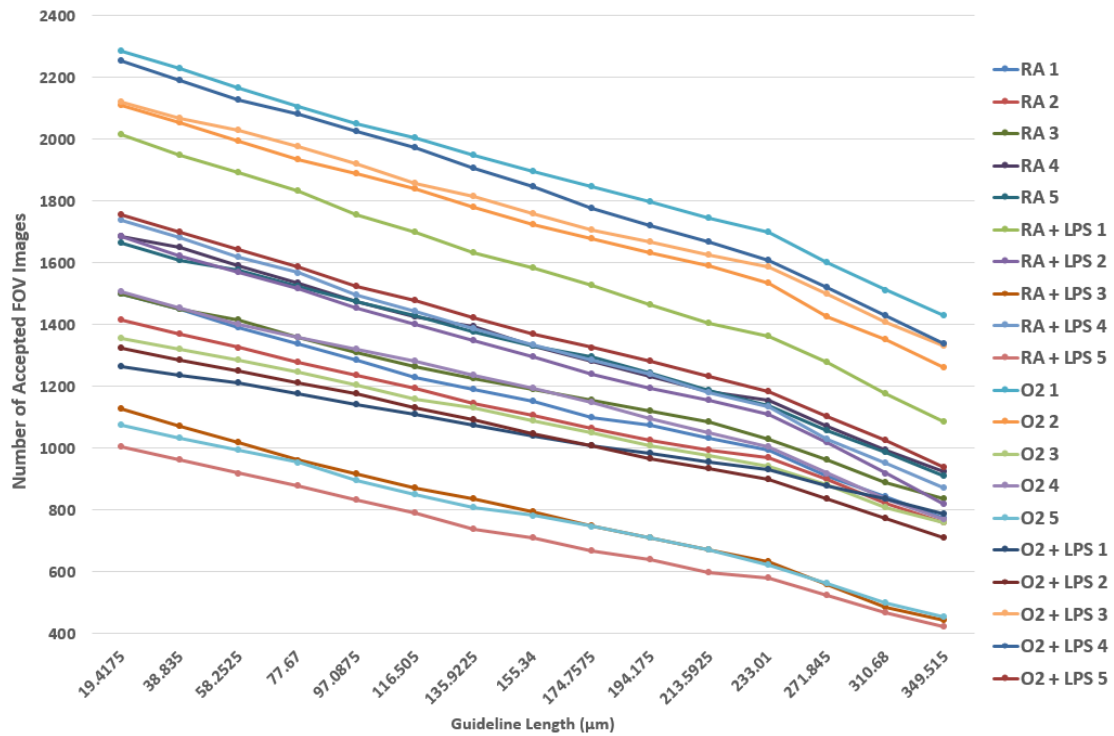


Figure 7.4 Number of accepted FOV images as a function of guideline length.

The number of accepted FOV images steadily decreased as the length of the superimposed guideline increased. The number of accepted FOV images varied from 2287 to 421 at a guideline length of 19.42 to 349.52 μm , respectively. Although the number of accepted FOV images decreases for larger guideline lengths, the lowest amount available for a single WSI to calculate the MLI score was 421, which is still above the recommended minimum. For each group (*i.e.*, RA, RA + LPS, O₂, O₂ + LPS), the resulting MLI scores, number of intersections, and number of chords are averaged, rather than displaying all 40 data lines in a single graph. To ensure that the number of accepted FOV images is not a factor in this analysis, the average MLI score for all groups

is repeated, but at a fixed number of 400 randomly selected accepted FOV images per WSI. The results are shown in Figure 7.5 for the indirect MLI method and Figure 7.6 for the direct method.

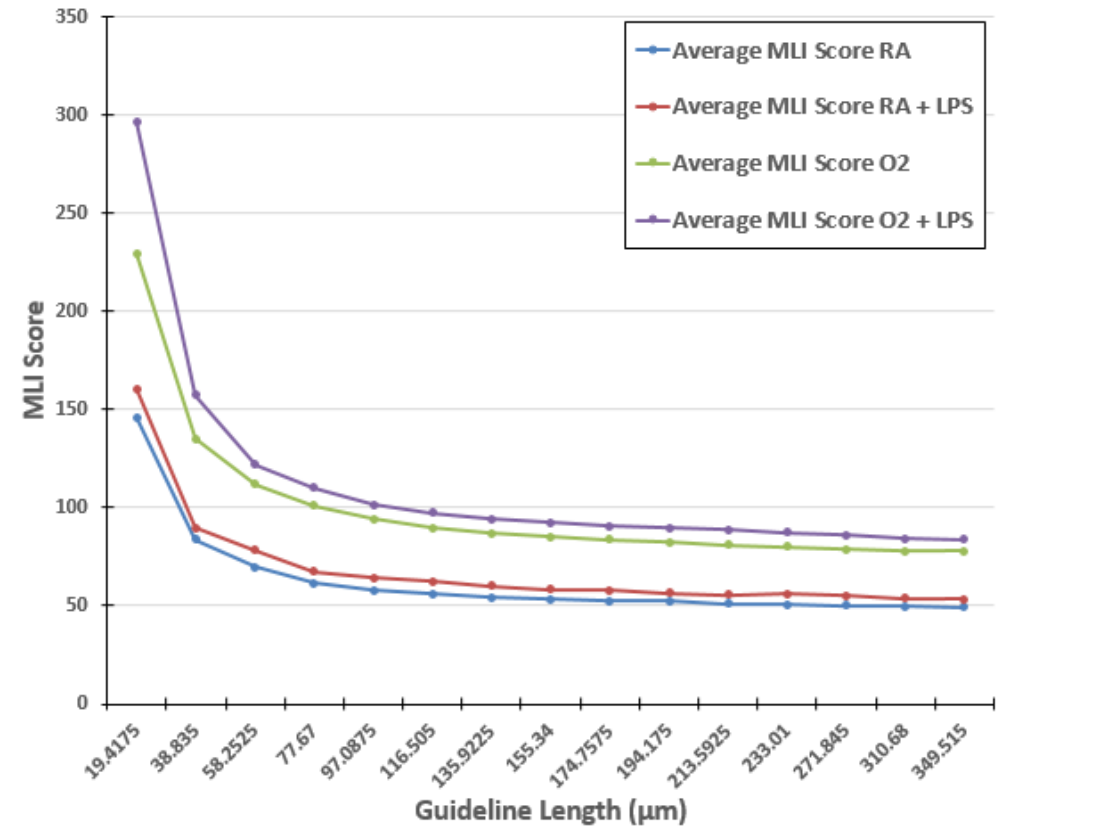


Figure 7.5 Average indirect MLI (MLI_{ind}) score per WSI group vs. guideline length (μm) at fixed number of accepted FOV images per WSI (400)

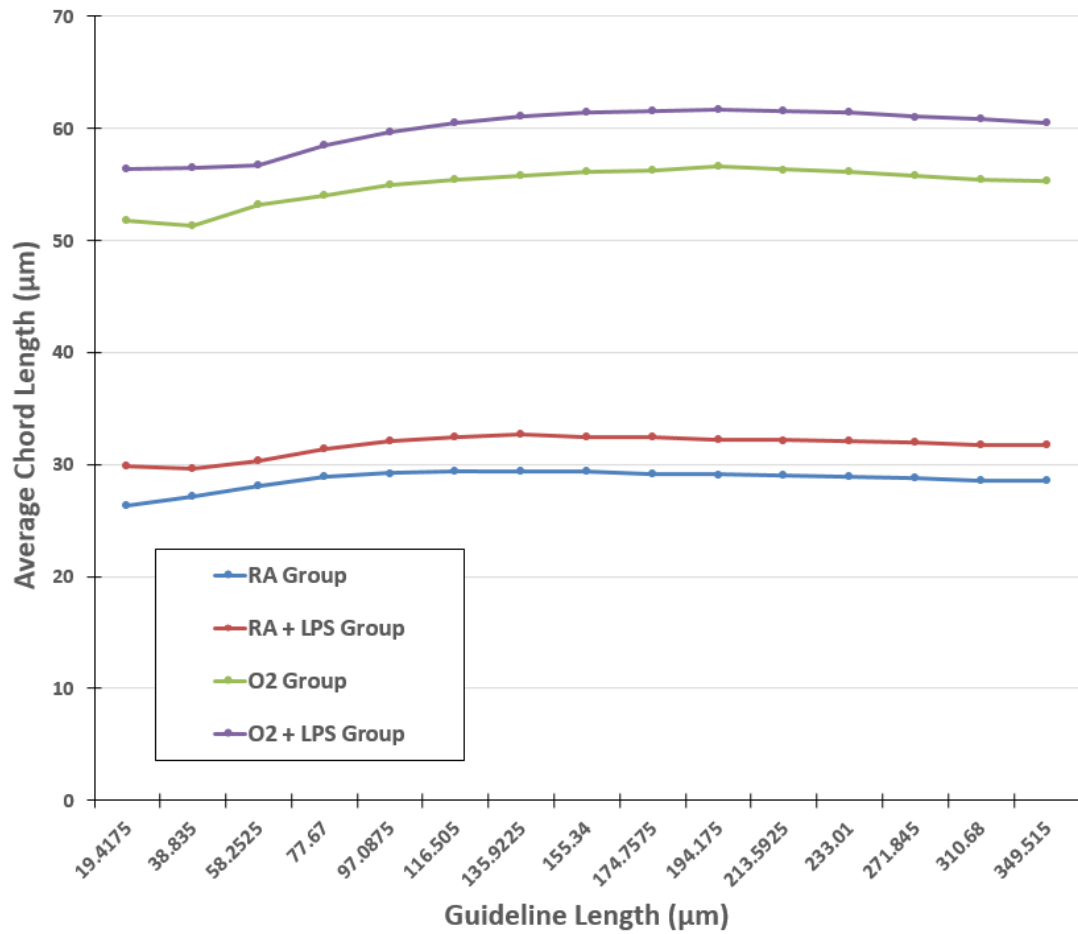


Figure 7.6 Average direct MLI (MLI_{dir}) score per WSI group vs. guideline length (μm) at fixed number of accepted FOV images per WSI (400)

For all WSI groups, repeated random sampling is performed to estimate the standard error associated with each guideline length. 400 randomly selected FOV images are used to calculate the MLI score for each WSI, with the results within each group averaged together. This is repeated 1000 times to get the standard error. The Results are shown in Table A2 of Appendix A. The results can be visualized in Figure 7.7, Figure 7.8, Figure 7.9, and Figure 7.10 which displays the average indirect and direct MLI score and standard error for the RA, RA + LPS, O₂, and O₂ + LPS group WSIs, respectively.

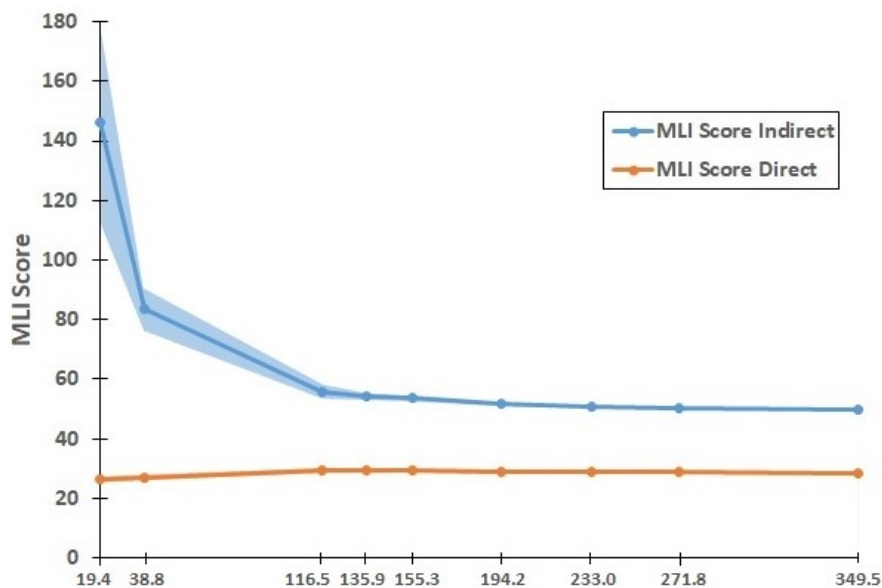


Figure 7.8 Average indirect MLI score (MLI_{ind}) and direct MLI score (MLI_{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for RA WSI Group

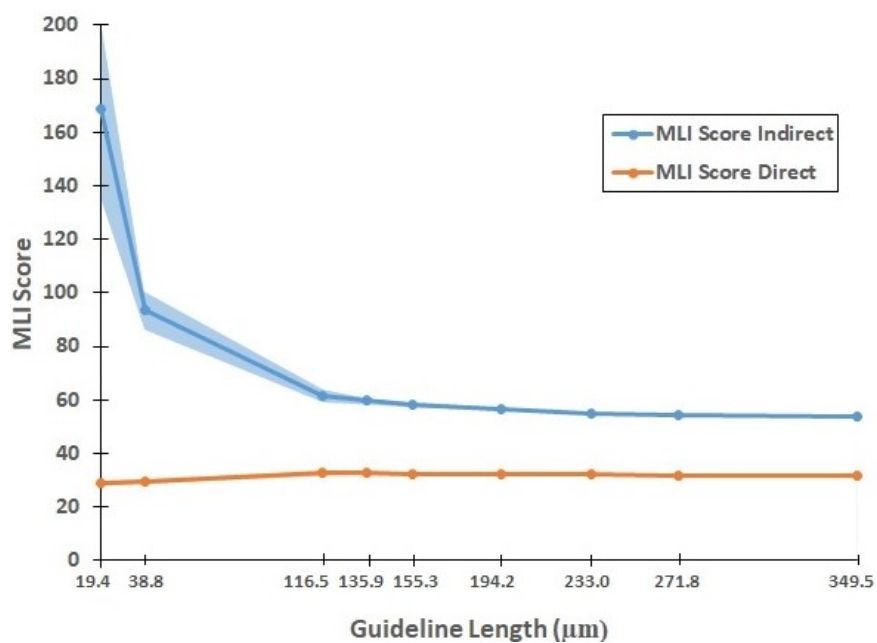


Figure 7.7 Average indirect MLI score (MLI_{ind}) and direct MLI score (MLI_{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for RA + LPS WSI Group

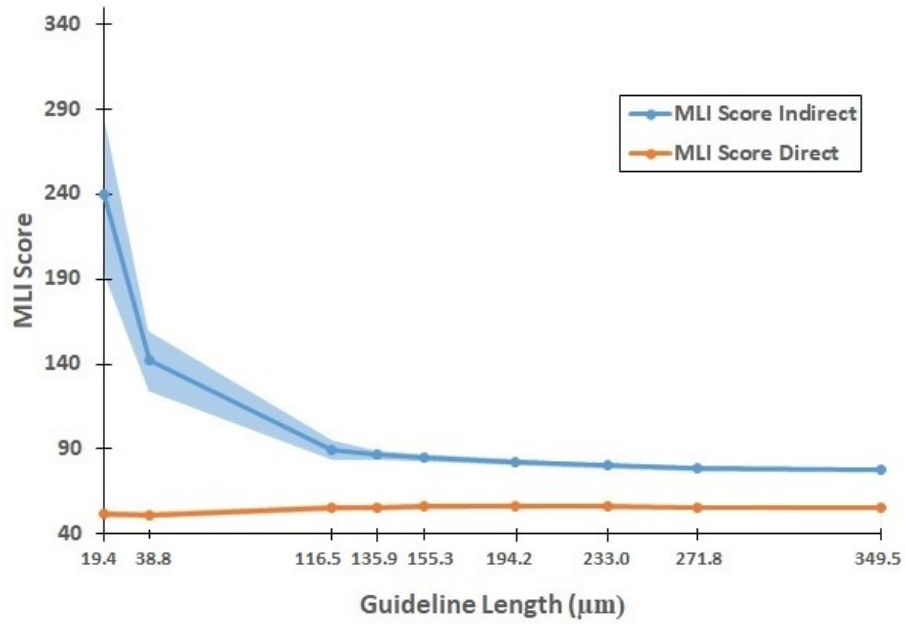


Figure 7.9 Average indirect MLI score (MLI_{ind}) and direct MLI score (MLI_{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for O₂ WSI Group

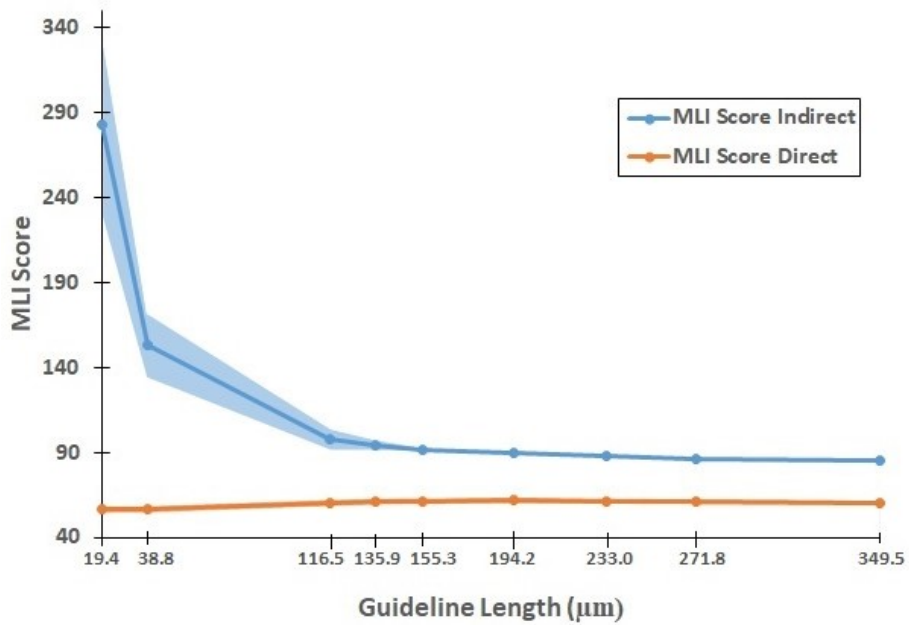


Figure 7.10 Average indirect MLI score (MLI_{ind}) and direct MLI score (MLI_{dir}) and associated standard error vs. guideline length (μm) at 400 accepted FOV images for O₂ + LPS WSI Group

7.4 Analyses and Discussion

As the guideline length increases, the number of accepted FOV images decreases somewhat linearly. This trend is expected, since an increase of guideline length increases the likelihood of contact between the guideline and a bronchi, blood vessel or pleural space. Therefore while the number of extracted FOV images per WSI remains the same for all guideline lengths, the number of accepted FOV images changes. The lowest number of accepted FOV images was 421, which is still greater than the recommended minimum of 300.

In Figure 7.2, as the guideline length increases, the MLI score decreases with a relationship similar to a decaying exponential. For very small guideline lengths, many FOV images that contain no intersections are accepted, thus resulting in a large increase in MLI score, as per Equation 2.1. Smaller guideline lengths also fail to have any intersections for larger alveoli, which in turn can largely increase the MLI score. This range of guideline lengths results in a non-linear relationship with the MLI score and a high standard error (*i.e.*, standard error of 7.22 and 18.79 for the RA and O₂ + LPS groups, respectively, at guideline length of 38.83 μm). As the guideline length increases further, its effect on the MLI score eventually starts to plateau (*i.e.*, standard error of 1.03 and 2.43 for the RA and O₂ + LPS groups, respectively, at guideline length of 155.34 μm). This plateauing begins approximately at a guideline length of 135.9 μm , where the standard error is approximately 1.26 and 3.03 for the RA and O₂ + LPS groups, respectively.

When observing the direct MLI score graph (Figure 7.3), there is almost no impact of guideline length on the resulting MLI score. Even for extremely small guideline lengths, the MLI score is relatively stable. These results are expected due to the nature of how the direct MLI score is calculated. If a guideline falls completely within an alveoli, which is typical for small guideline lengths and large alveoli, the edges of the guideline are extended until they reach an alveolar septa. Therefore typically at least a single chord is measured per FOV image, meaning that even with small guideline lengths, enough samples can be measured to accurately calculate the average chord length (direct MLI score). There is also a noticeably smaller standard error, even for small guideline lengths (*i.e.*, standard error of 0.92 and 1.98 for the RA and O₂ + LPS groups, respectively, at guideline length of 19.42 μm)

As shown in Figure 7.4, as guideline length increases, the number of accepted FOV images decreases somewhat linearly. This is expected, as an increase in guideline length increases the likelihood of contact between the guideline and a bronchi, blood vessel, or pleura. Therefore for smaller guideline lengths, there is a substantially higher number of accepted FOV images used for MLI calculation compared longer guideline lengths. To ensure that the number of accepted FOV images does not contribute to investigation, the number of accepted FOV images was fixed at 400, regardless of guideline length. As shown in Figure 7.5 and 7.6, the resulting MLI score follows the same pattern as before, even when fixed at 400 accepted FOV images.

The standard error at each guideline length was calculated through repeated random sampling at a fixed number of accepted FOV images for the RA, RA + LPS, O₂, and O₂ + LPS groups, shown in Figure 7.7, Figure 7.8, Figure 7.9, and Figure 7.10, respectively. For small guideline lengths, there is a noticeable standard error for the indirect MLI method (*e.g.*, guideline length of 19.42 μm had a standard error of 52.37 for the O₂ + LPS group) but a small standard error for the direct method (*e.g.*, guideline length of 19.42 μm had a standard error of 1.98 for the O₂ + LPS group). As guideline length increased however, the standard error for the indirect method largely decreased (*e.g.*, guideline length of 155.34 μm had a standard error of 2.31 for the O₂ + LPS group) while the standard error decreased slightly for the direct method (*e.g.*, guideline length of 155.34 μm had a standard error of 1.45 for the O₂ + LPS group). What qualifies as an acceptable standard error in MLI calculation is not defined in the relevant literature. In this work, the inter- and intra-rater variability that was found during human rater analysis from [37] was compared against the standard error associated with varying the guideline length. In future work, determining that threshold of acceptability regarding the standard error could help determine minimum requirements for guideline length when calculating the MLI score. The results suggest that the direct MLI method is less impacted by guideline length compared against the indirect MLI method, though the difference in impact is negligible for the guideline lengths typically used (*i.e.*, > 155.34 μm). The direct method also allows the user to obtain the variance of air space measures for the WSI and therefore a better understanding of the alveolar surface area. The results from this investigation suggest that the direct method for MLI calculation could be a more accurate and robust approach compared to indirect measurements, as it is less susceptible to variations caused by guideline length.

7.5 Conclusion

The fully automated system was used to assess how guideline length impacts the resulting MLI score for both indirect and direct methods. For indirect measurements of the MLI score, there is an inverse relationship between guideline length and MLI score; as the guideline length increased, the MLI score decreased like a decaying exponential. Using small guideline lengths (*i.e.*, $< 135 \mu\text{m}$) can result in inaccurate and misleading MLI calculations with a large standard error.

For direct measurements for MLI scoring, guideline length has little impact on the MLI score, suggesting that the direct method is perhaps a more robust approach to MLI scoring, with the further benefit of providing the variation of chord measurements of the lung and therefore a better understanding of the alveolar surface area. The time required to manually assess the MLI score by a human rater is on the order of hours, with the direct approach requiring even more time. The fully automated system however can perform both indirect and direct MLI scoring for a single WSI in 15 minutes. This analysis performed on guideline length would be extremely inefficient, costly and subject to human error if performed through human rater analysis, showing the versatility and research potential of the automated system.

Chapter 8: Conclusions and Future Work

8.1 Conclusions

The objectives of this thesis were the development of a modular and fully automated system to calculate the MLI score of digitized histopathology images of mouse lung tissue, an analysis into the differences between indirect and direct MLI scoring methods, and an investigation into the effects of guideline length and number of accepted FOV images on the resulting MLI score. The MLI score is a useful and extensively used method to assess air space size in histopathological images of lung tissue and is commonly used to monitor emphysematous changes to lung parenchyma. MLI scoring is predominantly performed in a semi-automated process, where FOV images are automatically presented to a human rater for analysis. Such methods, however, are inefficient, expensive, and susceptible to biases and errors. In this thesis, a modular pipeline was developed, that could fully automate the MLI scoring process of high-resolution WSIs of mouse lung tissue. The MLI score of the automated system was compared against the manual assessment of a human rater, showing a high degree of accuracy. The overall RMSD between the automated system and the human rater was 5.73 ± 5.66 , and a Pearson's correlation coefficient of $r=0.9931$. When compared against three human raters, the automated system displayed a higher degree of accuracy, with a RMSD between the automated system and the average of the three human raters of 4.71 ± 2.16 and a correlation of $r=0.9955$. These results suggest that the performance of the automated system might be better than the assessment of a single rater. The automated system could also differentiate between the different groups of the WSI dataset, showing a significant difference between the RA (M=53.40) vs. RA + LPS (M=59.85) groups ($U(5,5) = 1$, p-value = 0.016) and similarly a significant difference

between the O₂ (M=85.14) vs. O₂ + LPS (M=91.91) groups ($U(5,5) = 1$, p-value = 0.016), using the indirect MLI module. This demonstrates the potential of the automated system in differentiating between diseased and healthy lung tissue. While semi-automated evaluation can be on the order of hours per WSI, the processing time of the automated system was approximately 15 minutes, which could lead to a significant reduction in cost and increase in efficiency for researchers.

The modularity of the system allowed for enhancements or modifications to specific modules, without requiring changes to other modules of the pipeline. This was demonstrated through the use of two different modules in the *MLI Estimation* block; The *Intersection Counting* module, for indirect MLI scoring, and the *Chord Measurements* module, for direct MLI scoring. To investigate the differences between the indirect and direct methods for MLI scoring, the *MLI Estimation* module and the *Intersection Counting* module is utilized, respectively. The RMSD between the MLI scores of the two methods was 27.09 ± 0.71 , while displaying a high correlation ($r=0.9919$). By modifying the pipeline, the impact of *Septa Bias* and the *Partial Chord Bias* on the resulting MLI score was quantified.

By utilizing the fully automated system, the impact of guideline length and number of accepted FOV images on the resulting MLI score was quantified. For indirect MLI scoring, linear and non-linear regions were identified. A minimum number of accepted FOV images of approximately 300 FOV images had a standard error between 0.73 to 1.72 and 1.17 to 2.81 for direct and indirect methods, respectively, for the different mouse groups of the

dataset. Further increases in the number of accepted FOV images beyond 300 has a smaller impact on the standard error, but the standard error still decreases. With an automated system, using a large number of accepted FOV images (*e.g.*, 1000) is still advantageous, whereas it would be perhaps an unreasonable analyses involving human raters. These results suggest that using the direct method for MLI scoring is less susceptible to parameters such as guideline length, with a smaller standard error when compared to the indirect method. Such analysis would be extremely time-consuming if done through semi-automated methods, highlighting the benefit of the automated system.

8.2 Study Limitations

In this thesis, a dataset of 20 WSIs of mouse lung tissue was used, composed of four different groups. When analyzing the automated system ability to differentiate between groups, the analysis was performed on only five WSIs per group, and therefore the results could not be conclusive. While the results did suggest that the automated system was able to better differentiate between the different groups compared to the manual evaluation of a human rater, a much larger dataset is needed for conclusive results.

As displayed in Chapter 4, single rater evaluation could be subject to intra-rater variability and bias. To confirm the accuracy of the automated system, it would be beneficial to have more human raters perform MLI scoring on a larger WSI dataset to allow for more conclusive evidence and a better understanding of the automated systems performance.

8.3 Future Work

The work in this thesis has contributed to the automation of air space size assessment in WSIs of lung tissue, a greater understanding of how guideline length and number of accepted FOV images impact the MLI score, and identification of how two biases (i.e., *Septa Bias* and *Partial Chord Bias*) result in differences between direct and indirect MLI scoring methods. More work is needed to finalize the design of the system to allow for its use in a research setting. The subsections below provide a few recommendations of the work that can be conducted.

8.3.1 Development of a User-Friendly Graphical-User-Interface

The fully automated system was designed in Python (version 3.6.13) using Jupyter Notebook IDE during development and testing. To perform MLI scoring through the fully automated system, the user must provide the full folder path where the high-resolution WSIs are located, and run the resulting PY file. While the user can run the file on any IDE (i.e., Jupyter Notebook, Spyder... etc), it would be highly beneficial to design and implement a stand-alone desktop application. A desktop application where, for example, the user could drag and drop the WSIs they wish to analyze, would be beneficial for facilitating ease-of-use by researchers.

8.3.2 Investigation into analysis of WSIs of lung tissue for different species

The fully automated system utilized a deep learning model that was trained on the same dataset described in Section 4.2.1. It would be beneficial to evaluate the system and train the deep learning model on different animal species. Mouse lung tissue have major

difference to human lung tissue, with organ size being a significant differentiator. With larger airway diameter and alveolar size, it would be interesting to evaluate the generalizability of the fully automated system on different species. For example, rats are another common species used in lung research.

8.3.3 Addition of other methods for evaluating air space size and emphysematous changes

The MLI score is one of the most frequently used parameters for quantifying lung structure and emphysematous changes to lung parenchyma. There are, however, other parameters used for assessing lung structure, such as counting profiles per area (ex. Cells/FOV image or vessels/area), radial alveolar counts (RAC), and medial wall thickness [64]. With the binary segmentation masks generated in the semantic segmentation module and utilizing the general structure of the automated systems pipeline, it would be beneficial to implement other parameters for evaluating lung structure and compare the results against the relevant gold-standard.

8.3.4 Investigation into the effects of orientation and positioning of the superimposed guideline

The automated system was designed to calculate the MLI score in a way that mimicked human raters in the conventional semi-automated approach. In this method, a horizontal guideline is superimposed in the center of each FOV. The modularity of the automated system, however, can allow for modification of the guidelines positioning and orientation within the FOV image. This could increase the number of individual measurements that

contribute to the MLI score and may also reveal differences in the MLI score with regards to directionality and spatially within the lung.

Appendix A

Table A1. Average standard error at varying number accepted FOV images for each WSI group.

| Group | # Acc FOV | 10 | 50 | 100 | 200 | 300 | 400 | 1000 |
|-------|--------------|-------|------|------|------|------|------|------|
| RA | SE Dir | 4.25 | 1.84 | 1.30 | 0.89 | 0.73 | 0.61 | 0.34 |
| | SE Ind | 7.09 | 3.03 | 2.11 | 1.45 | 1.17 | 0.99 | 0.57 |
| RA + | SE Dir | 5.15 | 2.22 | 1.52 | 1.06 | 0.86 | 0.71 | 0.39 |
| LPS | SE Ind | 8.69 | 3.62 | 2.46 | 1.76 | 1.37 | 1.19 | 0.64 |
| O2 | SE Dir | 8.57 | 3.79 | 2.64 | 1.83 | 1.47 | 1.25 | 0.69 |
| | SE Ind | 14.89 | 5.91 | 4.25 | 2.98 | 2.37 | 2.07 | 1.14 |
| O2 + | SE Dir | 9.82 | 4.31 | 3.02 | 2.10 | 1.72 | 1.47 | 0.83 |
| LPS | SE Ind | 18.45 | 7.26 | 5.03 | 3.54 | 2.81 | 2.37 | 1.39 |

Table A1 provides a summary of the average standard error at varying number of accepted FOV images for each WSI group. The information from Table A1 is illustrated in Figure 6.1, Figure 6.2, Figure 6.3, and Figure 6.4 from Section 6.3

Table A2. Average standard error at varying guideline length for RA, RA + LPS, O₂, and O₂ +

LPS WSI group.

| Group | Guideline Length (μm) | 19.42 | 38.83 | 116.5 | 135.9 | 155.3 | 194.2 | 233 | 271.8 | 349.5 |
|----------------------|-----------------------|-------|-------|-------|-------|-------|-------|------|-------|-------|
| RA | SE Dir | 0.92 | 0.89 | 0.81 | 0.7 | 0.62 | 0.63 | 0.59 | 0.56 | 0.52 |
| | SE Ind | 33.59 | 7.22 | 2.52 | 1.26 | 1.03 | 0.97 | 0.91 | 0.84 | 0.79 |
| RA + LPS | SE Dir | 0.96 | 0.91 | 0.86 | 0.73 | 0.65 | 0.62 | 0.61 | 0.58 | 0.55 |
| | SE Ind | 37.63 | 9.53 | 3.48 | 1.85 | 1.23 | 1.15 | 1.04 | 0.98 | 0.92 |
| O ₂ | SE Dir | 1.93 | 1.82 | 1.77 | 1.69 | 1.27 | 1.19 | 1.1 | 1.04 | 0.97 |
| | SE Ind | 46.78 | 17.83 | 5.98 | 2.93 | 2.13 | 1.94 | 1.55 | 1.28 | 1.15 |
| O ₂ + LPS | SE Dir | 1.98 | 1.85 | 1.82 | 1.73 | 1.45 | 1.36 | 1.31 | 1.23 | 1.12 |
| | SE Ind | 52.37 | 18.79 | 6.17 | 3.42 | 2.31 | 1.92 | 1.86 | 1.76 | 1.68 |

Table A2 provides a summary of the average standard error at varying guideline lengths for each WSI group. The information from Table A2 is illustrated in Figure 7.7, Figure 7.8, Figure 7.9, and Figure 7.10 from Section 7.3.

References

- [1] World Health Organization. (2021, June. 21). *Chronic obstructive pulmonary disease (COPD)* [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- [2] National Heart, Lung, and Blood Institute. (2021, May. 12). Chronic obstructive pulmonary disease [Online]. Available: <https://www.nhlbi.nih.gov/health-topics/copd>
- [3] K. Yuki, M. Fujiogi, S. Koutsogiannaki., “COVID-19 pathophysiology: A review,” *Clin Immunol*, vol. 215, Jun 2020.
- [4] Q. Zhao, M. Meng, R. Kumar, Y. Wu, J. Huang, N. Lian, Y. Deng, S. Lin., “The impact of COPD and smoking history on the severity Of COVID-19: A systemic review and meta-analysis,” *J Med Virol*, vol. 92, issue. 10, pp. 1915-1921, Apr 2020.
- [5] D. D. Weiss, R. Casburi, R. Flannery, M. LeRoux-Williams, D. P. Tashkin., “A placebo-controlled, randomized trial of mesenchymal stem cells in COPD,” *Chest*, vol. 143, pp. 1590-1598, June 2013.
- [6] P. A. Kumar, Y. Hu, Y. Yamamoto, N. B. Hoe, T. S. Wei, D. Mu, et al., “Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection,” *Cell*. vol. 147, issue. 3, pp. 525-538, Oct 2011.
- [7] A. C. Inamdar, A. A. Inamdar, “Mesenchymal stem cell therapy in lung disorders: pathogenesis of lung diseases and mechanism of action of mesenchymal stem cell,” *Exp Lung Res*, vol. 39, issue. 8, pp. 315-327, Aug 2013.
- [8] J. W. Huh, S. Y. Kim, J. H. Lee, J. S. Lee, Q. V. Ta, M. Kim, et al., “Bone marrow cells repair cigarette smoke-induced emphysema in rats,” *Am J Physiol Lung Cell Mol Physiol*. Vol. 301, issue. 3, pp. L255-L266, Sep 2011.
- [9] P. Mehta, D. F. McAuley, M. Brown, E. Sanches, R. S. Tattersall, J. J. Manson., “COVID-19: consider cytokine syndromes and immunosuppression,” *Lancet*, vol. 395, issue. 10229, pp. 1033-1034, Mar 2020.
- [10] J. Chen, C. Hu, L. Chen, L. Tang, Y. hu, X. Xu, et al., “Clinical Study of Mesenchymal Stem Cell Treatment for Acute Respiratory Distress Syndrome Induced by Epidemic Influenza A (H7N9) Infection: A Hint for COVID-19 Treatment,” *Engineering*, vol. 6, issue. 10, pp. 1153-1161, Oct 2020.
- [11] C. C. W. Hsia, D. M. Hyde, M. Ochs, E. R. Weibel., “An Official Research Policy Statement of the American Thoracic Society/European Respiratory Society: Standards for Quantitative Assessment of Lung Structure,” *Am J Respir Crit Care Med*, vol. 181, issue. 4, pp. 394-418, Feb 2010.
- [12] J. H. Lee, D. S. Lee, E. K. Kim, K. H. Choe, Y. M. Oh, T. S. Shim, et al., “Simvastatin inhibits cigarette smoking-induced emphysema and pulmonary hypertension in rat lungs,” *Am J Respir Crit Care Med*, vol. 172, issue. 8, pp. 987-993, Oct 2005.
- [13] E. J. Caraher, S. Kwon, S. H. Haider, G. Crowley, A. Lee, M. Ebrahim, et al., “Receptor for advanced glycation end-products and World Trade Center particulate induced lung function loss: a case-cohort study and murine model of acute particulate exposure,” *Plos One*, vol. 19, issue. 9, Sep 2017.

- [14] H. Parameswaran, A. Majumdar, S. Ito, A. M. Alencar, B. Suki., “Quantitative characterization of airspace enlargement in emphysema,” *J Appl Physiol*, vol. 100, issue. 1, pp. 186-193, Jan 2006.
- [15] S. H. Poggi, C. Salafia, S. Paiva, N. J. Leak, J. C. Pezzullo, A. Ghidini., “Variability in pathologists’ detection of placental meconium uptake,” *Am J Perinatol*. vol. 26, issue. 3, pp. 207-210, Mar 2009.
- [16] L. Mulrane, E. Rexhepaj, S. Penney, J. J. Callanan, W. M. Gallagher, “Automated image analysis in histopathology: a valuable tool in medical diagnostics,” *Expert Rev Mol Diagn*, vol. 8, issue. 6, pp. 707-725, Nov 2008.
- [17] J. D. Webster, A. M. Michalowski, J. E. Dwyer, K. N. Corps, B. R. Wei, T. Juipperi, et al., “Investigation into diagnostic agreement using automated computer-assisted histopathology pattern recognition image analysis,” *J Pathol Inform*, vol. 3, issue 18, Apr 2012
- [18] C. Bouzin, M. L. Saini, K. K. Khaing, J. Ambroise, E. Marbaix, V. Gregoire, V. Bol., “Digital pathology: elementary, rapid and reliable automated image analysis,” *Histopathology*, vol. 69, issue. 6, pp. 888-896, Sep 2015.
- [19] I. R. I. Haque, J. Neubert., “Deep learning approaches to biomedical image segmentation,” *Infor in Med Unl*, vol. 18, Jan 2020.
- [20] C. C. W. Hsia, D. M. Hyde, E. R. Weibel, “Lung Structure and the Intrinsic Challenges of Gas Exchange,” *Compr Physiol*, vol. 6, issue. 2, pp. 827-895, Mar 2016.
- [21] R. L. Riley, A. Cournand, “Ideal Alveolar Air and the Analysis of Ventilation-Perfusion Relationships in the Lungs,” *J Appl Physiol*, vol. 1, issue. 12, pp. 825-847, Jun 1949.
- [22] Wikimedia Foundation. (2022, May 6). *Lung*. *Wikipedia* [online] Available: <https://en.wikipedia.org/wiki/Lung>
- [23] M. H. Ross, W. Pawlina, *Histology*, 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2002
- [24] B. Tielemans, K. Dekoster, S. E. Verlede, S. Sawal, B. Leszczyński, K. Laperre, et al., “From Mouse to Man and Back: Closing the Correlation Gap between Imaging and Histopathology for Lung Diseases,” *Diagnostics*, vol. 10, issue. 9, pp. 636, Aug 2020.
- [25] A. H. Fischer, K. A. Jacobson, J. Rose, R. Zeller., “Hematoxylin and Eosin Staining of Tissue and Cell Sections,” *CSH Protoc*, May 2008.
- [26] C. J. L Murray, A. D. Lopez, *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*, vol. 1, Cambridge, MA: Harvard University Press, 1996.
- [27] P. M. Henson, G. P. Cosgrove, R. W. Vandivier, “Apoptosis and Cell Homeostasis in Chronic Obstructive Pulmonary Disease,” *Proc Am Thorac Soc*. vol. 3, issue. 6, pp. 512-516, Aug 2006.
- [28] T. Dey, J. Kalita, S. Weldon, C. C. Taggart, “Proteases and Their Inhibitors in Chronic Obstructive Pulmonary Disease,” *J Clin Med*. vol. 7, issue. 9, Aug 2018.
- [29] W. Mitzner, “Use of mean airspace chord length to assess emphysema,” *J Appl Physiol*, vol. 105, issue. 6, pp. 1980-1981, Dec 2008.
- [30] F. S. Rosenthal, Z. A. Begum, “Image-based determination of chord lengths in air-dried lungs,” *J Microsc*, vol. 219, issue. 3, pp. 160-166, Sep 2005.

- [31] S. Brarber, K. A. T. Verheijden, P.A. Henricks, A.D. Kraneveld, G. Folkerts, "A comparison of fixation methods on lung morphology in a murine model of emphysema," *American Journal of Physiology*, vol. 299, issue. 6, pp. L843-L851, Dec 2010.
- [32] X. Yan, J.J.P. Carbayo, E. R. Weibel, C. C. W. Hsia, "Variation of lung volume after fixation when measured by immersion or Cavalieri method," *Am J Physiol Lung Cell Mol Physiol*, vol. 284, issue. 1, pp. L242-L245, Sep 2002.
- [33] T. H. van Kuppevelt, A. A. Robbesom, E. M. Versteeg, J. E. Veerkamp, C. L. Herwaarden, P. N. Dekhuijzen, "Restoration by vacuum inflation of original alveolar dimensions in small human lung specimens," *Eur Respir J*, vol. 15, issue. 4, pp. 771-777, Apr 2000.
- [34] S. E. Soutiere, W. Mitzner, "On defining total lung capacity in the mouse," *J Appl Physiol*, vol 96, issue. 5, pp. 1658-1664. May 2004.
- [35] Y. Horai, T. Kakimoto, K. Takemoto, M. Tanaka, "Quantitative analysis of histopathological findings using image processing software," *J Toxicol Pathol*, vol. 30, pp. 351-358, Oct 2017.
- [36] G. Crowley, S. Kwon, E. J. Caraher, S. H. Haider, R. Lam, P. Batra, et al., "Quantitative lung morphology: semiautomated measurement of mean linear intercept," *BMC Pulm Med*, vol. 19, issue. 1, pp. 206-209, Nov 2019.
- [37] S. Salsabili, M. Lithopoulos, S. Sreeraman, A. Vadivel, B. Thebaud, A. D. C. Chan, E. Ukwatta., "Fully automated estimation of the mean linear intercept in histopathology images of mouse lung tissue," *J Med Imaging*, vol. 8, issue. 2, Mar 2021.
- [38] Knudsen L, Weibel ER, Gundersen HJG, Weinstein F V., Ochs M. Assessment of air space size characteristics by intercept (chord) measurement: An accurate and efficient stereological approach. *J Appl Physiol*. 2010. doi:10.1152/japplphysiol.01100.2009
- [39] C. Sallon et al., "Automated high-performance analysis of lung morphometry," *Am. J. Respir. Cell Mol. Biol*, vol. 53, issue. 2, pp. 149-158, Aug 2015.
- [40] E. Rieger-Fackeldey, M. S. Park, B. L. Schanbacher, M. S. Joshi, L. G. Chicoine, L. D. Nelin, et al., "Lung development alterations in newborn mice after recovery from exposure to sublethal hyperoxia," *Am J Pathol*, vol. 184, issue. 4, Apr 2014.
- [41] A. A. Robbesom, E. M. M. Versteeg, J. H. Veerkamp, J. Krieken, H. Bulten, H. Smits, et al. "Morphological Quantification of Emphysema in Small Human Lung Specimens: Comparison of Methods and Relation with Clinical Data," *Mod Pathol*, vol. 16, issue. 1, pp. 1-7, Jan 2003.
- [42] W. M. Thurlbeck, "Internal surface area and other measurements in emphysema," *Thorax*. vol. 22, issue. 6, pp. 483-496, Nov 1967.
- [43] E. Reinhard, M. Ashikhmin, B. Gooch, P. Shirley, "Color transfer between images," *IEEE Comput Graph Appl*, vol. 21, issue. 5, pp. 34-41, Oct 2001.
- [44] B. Ao, X. Wu, Q. Wen, K. Li, J. Yin., "Segmentation of renal cell carcinoma using stain normalization," in *2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Guangzhou, China, 2022, pp. 146-149.
- [45] J. He, X. Jia, S. Chen, J. Liu, "Multi-Source Domain Adaptation with Collaborative Learning for Semantic Segmentation", in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Virtual, 2021, pp. 11003-11012.

- [46] M. Owais, N. R. Baek, K. R. Park, "DMDF-Net: Dual Multiscale Dilated Fusion Network for Accurate Segmentation of Lesions Related to COVID-19 in Lung Radiographic Scans", *Expert Syst Appl*, vol. 202, May 2022.
- [47] C. Li, H. Chen, X. Li, N. Xu, Z. Hu, D. Xue, et al., "A review for cervical histopathology image analysis using machine vision approaches," *Artif Intell Rev*, vol. 53, pp. 4821-4862, Feb 2020.
- [48] X. Zhou et al., "A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks," *IEEE Access*, vol. 8, May 2020.
- [49] S. Salsabili, E. Ukwatta, A. D. C. Chan., "Automated Segmentation of Complex Biological Structures in Histopathology Images," Ph.D. dissertation, Dept. Syst. Comp. Eng., Carleton University, Ottawa, ON, 2022
- [50] J. M. Bland, D. G. Altman., "Measuring agreement in method comparison studies," *StatMethods Med Res.* vol. 2, pp. 135-160, Jun 1999.
- [51] M. Hurskainen, I. Mižíková, D. P. Cook, N. Andersson, C. Cyr-Depauw, F. Lesage, et al., "Single cell transcriptomic analysis of murine lung development on hyperoxia-induced damage," *Nat Commun*, vol. 12, issue. 1, Mar 2021.
- [52] A. Vadivel, R. S. Alphonse, L. Ionescu, D. S. Machado, M. O'Reilly, F. Eaton, et al., "Exogenous Hydrogen Sulfide (H₂S) Protects Alveolar Growth in Experimental O₂-Induced Neonatal Lung Injury," *PLoS One*, vol. 9, issue. 3, Mar 2014.
- [53] R. Bals., "Lipopolysaccharide and the lung: a story of love and hate," *Eur Resp J*, vol. 25, issue. 5, pp. 776-777, May 2005.
- [54] C. Barazzone, S. Horowitz, Y.R. Donati, I. Rodrigues, P.F. Piguet., "Oxygen toxicity in mouse lung: pathways to cell death," *Am J Respir Cell Mol Biol.* vol. 19, issue. 4, pp. 573-581, Oct 1998.
- [55] B. Rosner, D. Grove., "Use of the Mann-Whitney U-test for clustered data," *Stat Med.* vol. 18, issue. 11, pp. 1387-1400, Jun 1999.
- [56] H. J. Jacob., "Functional Genomics and Rat Models," *Genome Res*, vol. 11, pp. 1013-1016, Nov 1999.
- [57] B. Charreau, L. Tesson, J. P. Soulillou, C. Pourcel, I. Anegon, "Transgenesis in rats: Technical aspects and models," *Transgenic Res.* vol. 5, issue. 4, pp. 223-234, Jul 1996.
- [58] L. Birke, "Who—or what—are the rats (and mice) in the laboratory," *Soc Anim.*, vol. 13, issue. 3, pp. 207-224, Feb 2003.
- [59] R. K., Albert, W. Kirk, C. Pitts, J Butler., "Extra-alveolar vessel fluid filtration coefficients in excised and in situ canine lobes," *J Appl Physiol*, vol. 59, issue. 5, pp. 1555-1559, Nov 1985.
- [60] P. L. Khimenko, A. E. Taylor., "Segmental microvascular permeability in ischemia reperfusion injury in rat lung," *Am J Physiol Lung Cell Mol*, vol. 276, issue. 6, pp. L958-L960, Jun 1999.
- [61] W. Truog, D. Xu, I. Ekekezie, S. Mabry, M. Rezaiekhalthigh, S. Svojanovsky, M. J. Soares, "Chronic Hypoxia and Rat Lung Development: Analysis by Morphometry and Directed Microarray," *Pediatr Res*, vol. 64, issue. 1, pp. 56-62, Jul 2008.
- [62] H. Sahebjami, J. A. Wirman., "Emphysema-like Changes in the Lungs of Starved Rats," *Am Rev Respir Dis*, vol. 124, issue. 5, Mar 1981.

- [63] J. C. Parker, M. I. Townsley, "Evaluation of lung injury in rats and mice," *Am J Physiol-Lung Cell Mol Physiol*, vol. 286, issue. 2, pp. L231-L246, Feb 2004.
- [64] C. Mühlfeld, J. Hegermann, C. Wrede, Matthias Ochs., "A review of recent developments and applications of morphometry/stereology in lung research," *Amer Physiol Soc*, vol. 309, Issue. 6, pp. L526-L536, Sep 2015.
- [65] M. Saetta, R. J. Shiner, G. E. Angues, W. D. Kim, N. S. Wang, M. King, *et al.*, "Destructive index: a measurement of lung parenchymal destruction in smokers," *Am Rev Respir Dis*, vol. 131, issue. 5, pp. 764-769, May 1985.
- [66] A. Nagai, I. Yamawaki, W. M. Thurlbeck, T. Takizawa, "Assesment of lung parenchymal destruction by using routine histologic tissue sections," *Am Rev Respir Dis*, vol. 139, issue. 2, pp. 313-319, Feb 1989.
- [67] K. J. Myall, B. Mukherjee, A. M. Castanheira, J. L. Lam, G. Benedetti, S. M. Mak *et al.*, "Persistent Post-COVID-19 Interstitial Lung Disease. An Observational Study of Corticosteroid Treatment," *Ann Am Thorac Soc*, vol. 18, issue. 5, pp. 799-806, May 2021.
- [68] T. S. Elhakim, H. S. Abdul, C. P. Romero, Y. R. Fuentes, "Spontaneous pneumomediastinum, pneumothorax and subcutaneous emphysema in COVID-19 pneumonia: a rare case and literature review," *BMJ Case Rep*, vol. 13, issue. 12, Dec 2020.
- [69] S. C. Lee, K. J. Sun, C. H. Han, S. C. Park, J. Y. Jung, "Impact of COPD on COVID-19 prognosis: A nationwide population-based study in South Korea," *Sci Rep*, vol. 1, issue. 3735, Feb 2021.