CHARACTERIZATION OF PERIPHERAL LUNG LESIONS BY STATISTICAL IMAGE PROCESSING OF ENDOBRONCHIAL ULTRASOUND IMAGES

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science Biomedical Engineering

By

Aaron T. Madaris
B.S.B.M.E, Wright State University, 2015

2016
Wright State University
I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Aaron T. Madaris ENTITLED Characterization of Peripheral Lung Lesions by Statistical Image Processing of Endobronchial Ultrasound Images BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Biomedical Engineering.

Ulas Sunar, Ph.D.
Thesis Director

Jaime E. Ramirez-Vick, Ph.D.
Chair, Department of Biomedical, Industrial and Human Factors Engineering

Committee on Final Examination

Jaime Ramirez-Vick, Ph.D

Jason Parker, Ph.D

Ulas Sunar, Ph.D

Robert E. W. Fyffe, Ph.D.
Vice President for Research and Dean of the Graduate School
Abstract


This thesis introduces the concept of implementing greyscale analysis, also known as intensity analysis, on endobronchial ultrasound (EBUS) images for the purposes of diagnosing peripheral lung tumors. The statistical methodology of using greyscale and histogram analysis allows the characterization of lung tissue in EBUS images. Regions of interest (ROI) will be analyzed in MATLAB and a feature vector will be created. A feature vector of first-order, second-order and histogram greyscale analysis will be created and used for the classification of malignant vs benign peripheral lung tumors. The tools that were implemented were MedCalc for the initial statistical analysis of receiver operating curves (ROC), Multiple Regression and MATLAB for the machine learning and ROI collection. Feature analysis, multiple regression and machine learning methods were used to better classify the malignant and benign EBUS images. The classification is assessed with a confusion matrix, ROC curve, accuracy, sensitivity and specificity. It was found that minimum pixel value, contrast and energy are the best determining factors to discriminate between benign and malignant EBUS images.
# TABLE OF CONTENTS

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

1.1 Research Motivation ............................................................................................. 1

1.2 Research Objective ................................................................................................ 4

2 METHODS .................................................................................................................. 6

2.1 Endobronchial Ultrasound Imaging of Lung Tissue ............................................ 6

2.2 Grey scale Analysis ................................................................................................. 8

2.3 Quantitative Parameters of Grey Scale Analysis ............................................... 10

2.4 Receiver Operating Curve Analysis ..................................................................... 17

2.5 Multiple Regression ............................................................................................... 20

2.6 Machine Learning ................................................................................................. 22

2.6.1 Support vector machine (SVM) Method .......................................................... 25

2.7 EBUS ROI selection ............................................................................................ 28

2.7.1 Whole lesion ROI ............................................................................................ 28

2.7.2 Multiple ROI ................................................................................................. 28

2.8 Machine Learning setup ..................................................................................... 29

2.8.1 Constructing the Feature Vector .................................................................... 29

2.8.2 Cross Validation and Hold-Out Methods ....................................................... 30

2.8.3 Machine Selection .......................................................................................... 30

2.8.4 Training and exporting the Machine ............................................................... 31

3 RESULTS .................................................................................................................. 32

3.1 ROI Selection of the EBUS Images ..................................................................... 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 MedCalc Analysis of the Image Parameters (N=13)</td>
<td>36</td>
</tr>
<tr>
<td>3.3 MedCalc Analysis of the Image Parameters (N=130)</td>
<td>39</td>
</tr>
<tr>
<td>3.4 Multiple Regression Method (N=130-samples)</td>
<td>41</td>
</tr>
<tr>
<td>3.5 MATLAB Machine Learning on the ROI datasets</td>
<td>44</td>
</tr>
<tr>
<td>3.6 Comparison of the N=13 to N=130 Methods</td>
<td>54</td>
</tr>
<tr>
<td>4 CONCLUSION AND FUTURE DIRECTIONS</td>
<td>58</td>
</tr>
<tr>
<td>Bibliography</td>
<td>60</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1 Lung Cancer 5 year survival rate................................................................. 1
Figure 1.2 EBUS procedure ....................................................................................... 2
Figure 2.1 Transbronchial needle with convex probe transducer .............................. 7
Figure 2.2 Human vision vs Computer vision of the grey scale ............................... 8
Figure 2.3 Analysis scheme ..................................................................................... 9
Figure 2.4 Pixel selection and Histogram analysis .................................................... 11
Figure 2.5 Histogram features of Skewness and Kurtosis ........................................ 12
Figure 2.6 How the Computer sees 5 different textural patterns ............................ 13
Figure 2.7 Contrast example .................................................................................... 14
Figure 2.8 Low Energy vs High Energy ................................................................. 15
Figure 2.9 Low Homogeneity vs High Homogeneity .............................................. 16
Figure 2.10 Different types of ROC curves and how they assess the model .......... 17
Figure 2.11 The Rankings of the classifiers based on the Area Under the Curve ...... 18
Figure 2.12 Cutoff points ....................................................................................... 18
Figure 2.13 Multiple Regression Example ............................................................. 22
Figure 2.14 The goal of Machine Learning ............................................................. 22
Figure 2.15 Linear SVM Example ......................................................................... 26
Figure 2.16 Cubic SVM and Radial SVM .............................................................. 27
Figure 3.1 Malignant EBUS image whole lesion method ........................................ 32
Figure 3.2 Benign EBUS image whole lesion method ........................................... 33
Figure 3.3 Malignant EBUS image multiple ROI method .................................... 34
Figure 3.4 Benign image multiple ROI method..................................................... 35
Figure 3.5 ROC for Homogeneity, Energy, Contrast and Minimum whole lesion .... 38
Figure 3.6 ROC Homogeneity, Energy, Contrast and Minimum multiple ROI ...... 40
Figure 3.7 The ROC curves of the multiple regression combinations .................. 42
Figure 3.8 Energy vs Contrast scatter plot ................................................................. 44
Figure 3.9 SVM ROC of the Min Contrast and Energy full lesion method .............. 45
Figure 3.10 Confusion Matrix of the Min Contrast and Energy full lesion method .... 45
Figure 3.11 SVM ROC of the Min Contrast and Homogeneity full lesion method .... 46
Figure 3.12 Confusion Matrix Min Contrast and Homogeneity full lesion method .... 47
Figure 3.13 Confusion Matrix Contrast combined with Energy 130-sample method .... 48
Figure 3.14 Energy and Contrast 130 sample ROC .................................................... 48
Figure 3.15 Confusion Matrix Contrast Min Energy 130-sample method ................. 49
Figure 3.16 ROC of Min Contrast and Energy 130 samples ...................................... 49
Figure 3.17 Confusion Matrix Min Contrast Homogeneity 130-sample method ........ 50
Figure 3.18 ROC of Min, Contrast, Homogeneity 130 sample ............................... 50
Figure 3.19 Machine Learning, Multiple Regression and Histogram/Texture Rankings 52
List of Tables

Table 2.1 Confusion Matrix.................................................................................................................. 19
Table 2.2 Correlation Matrix .................................................................................................................. 21
Table 3.1 13 sample MedCalc Results................................................................................................. 36
Table 3.2 130 sample MedCalc Results................................................................................................. 39
Table 3.3 Combined features with Cut point, Sensitivity, Specificity and Youden Index 43
Table 3.4 Combined features with LR, AUC and p-value................................................................. 43
Table 3.5 SVM Results ....................................................................................................................... 51
Table 3.6 SVM vs Multiple Regression Results.................................................................................. 53
Table 3.7 Malignant t-test .................................................................................................................. 55
Table 3.8 Benign t-test ......................................................................................................................... 56
Acknowledgments

I would like to take this opportunity to extend my thanks to my committee members, Drs. Ulas Sunar, Jaime Ramirez-Vick and Jason Parker, who have shown me the qualities and the drive that it takes to be a Biomedical Engineer. I would like to also thank Kassem Harris, MD from Roswell Park Cancer Institute for providing ultrasound images.
Dedicated to

My fiancé Tori Turner who has always been there for me when I needed her. Who always believed in the work that I was completing and to allow me the opportunity to continue my schooling.
1 INTRODUCTION

1.1 Research Motivation

Non-Small-Cell Lung Carcinoma (NSCLC) is a disease that can lead to death quickly. The sub-categories of NSCLC are: Adenocarcinoma, squamous-cell carcinoma and small-cell carcinoma. Adenocarcinoma and squamous-cell carcinoma make up almost 40% and 30% of all the lung cancer, respectively [1]. It is vital to diagnose NSCLC early to prevent its spread and improve survival (Figure 1.1). In this thesis research, Endobronchial Ultrasound (EBUS) greyscale analysis is implemented to identify malignant and benign lymph node tumors in the distal regions of the lung in early stage.

![Figure 1.1 Lung Cancer 5 year survival rate](image)

*Figure 1.1 Lung Cancer 5 year survival rate*
EBUS as seen in Figure 1.2 is a combination of an endoscope and ultrasound. It is fed through the trachea into the peripheries of the lung and the imaging device is pressed up against the bronchial wall for imaging. Physicians currently use this approach during screening to identify suspicious areas to take biopsies. Under imaging guidance biopsies are performed with a transbronchial needle that is part of endoscope. Figure 1.2 shows the biopsy and ultrasound imaging capabilities of the EBUS and as it traverses down the trachea to the desired lymph node.

![Figure 1.2 EBUS procedure. Retrieved from [2]](image)

Clinicians perform multiple biopsies to characterize lung tissue, and image processing can help guide clinicians during biopsy procedures by using quantitative imaging metrics. General practice for obtaining a biopsy is to randomly select in the region of interest. The biopsy samples are then sent to a pathologist for histological analysis as a gold standard. EBUS procedures have been performed with a convex probe and a mini probe to predict the metastatic lymph nodes [3–14]. The EBUS with a convex probe is capable of producing B-mode greyscale images as well as Doppler color images in real time. The greyscale images show structural and tonal features of the underlying
Mediastinal lymph node metastasis has the features of distinct margin, heterogeneous echogenicity and the presence of necrosis signatures [3]. The goal of characterizing these features is to analyze individual pixel value as well as texture as discriminating factors between malignant and benign lesions. Although the morphological features in EBUS can provide contrasts for the identification of tumors, statistical image processing-based approaches studied in this thesis is expected to improve discrimination of malignant from benign tissue by providing quantitative imaging measures. Greyscale histogram and texture analysis has been applied previously ultrasound imaging for characterization of breast, liver, prostate and other tumor types [3, 4, 11–13, 15–19]. Raw, gray-scale ultrasound images may be hard to interpret by the clinician during the biopsy procedure; because, the images may show low contrast, and the contrast itself might not sufficient enough for accurate diagnosis. Thus, biopsy sampling is mostly random and the yield can be as low as 30% for the detection of peripheral lung lesions. Moreover, the biopsy sampling can induce risks to patients, time-consuming and can be costly. The outcome of the biopsy analysis can take up to 1 week. Earlier diagnosis, preferably at the intraoperative settings, can provide earlier interventions, which can be crucial for the clinical outcome. Therefore, quantitative image processing analysis can provide more accurate characterization of samples, which could lead to biopsies obtained from the most relevant region of interest. Ultimately, it can guide clinicians at the intraoperative settings, reduce randomness of the clinical biopsy sampling, help histo-pathologists classify the tissue faster and decrease patient recovery time.
1.2 Research Objective

The main objective of this project is to provide a quantitative approach for accurate diagnosis of peripheral lung tumors by using statistical image processing methods of endobronchial ultrasound (EBUS) images. More specifically, to identify image features that classify tumor as either malignant or benign. By identifying image features early there will be a reduction in the time required for the histo-pathologist to classify the tissue, a reduction in the number of biopsies needed to make the diagnosis with possible less complications such as bleeding and pneumothorax and a reduction in patient recovery time and the overall procedure will become faster for early intervention and ultimately for improved clinical outcome.

EBUS is being used in the clinic at the intraoperative settings to help guide biopsy sampling [3–14, 20, 21]. The main purpose of EBUS during biopsy sampling is to narrow down an area of interest to be biopsied. During image-guided biopsy sampling procedure, EBUS transbronchial needle aspiration (TBNA) is used to obtain biopsies while obtaining EBUS images. The biopsied samples are analyzed by cytology or histology, the current gold standard for tissue diagnosis.

Imaging techniques such as greyscale analysis that involves texture analysis and histogram analysis are used to classify the EBUS images, which are then compared with the histology results. The reports utilizing greyscale analysis have shown substantial variations, which could be due to instrumentation as well as variations in image analysis. Overall, in this thesis work, statistical imaging approaches are implemented for the optimal classification of EBUS images [22]. The hypothesis is that raw ultrasound images do not have provide sufficient contrast during visualization at the intraoperative settings,
and that imaging derived quantitative parameters will allow additional discrimination power between malignant and benign tissue samples.
2 METHODS

2.1 Endobronchial Ultrasound Imaging of Lung Tissue

Endobronchial ultrasound (EBUS) is currently being used to identify the location of tumors in the peripheries of the lung. EBUS can use a convex curvilinear multi-element probe transducer or a rotating single element radial probe transducer (also called radial EBUS) to send (pulse) and receive (echo) sound waves to obtain ultrasound images. The output from the echo signal is converted into an image that can be used for diagnostic assessment. The previous studies have used B-mode EBUS images [3–14, 20, 21], where B-mode stands for the brightness mode of an ultrasound image. It is based on intensity values that can be interpreted in a grey scale image, which can then be analyzed statistically via grey scale analysis. The EBUS miniaturized probe is used to reach the peripheries of the lung for obtaining high quality, high-resolution EBUS images. EBUS with a miniaturized probe has been implemented in several studies [11, 12, 14, 20]. The use of the miniaturized probe has increased the accuracy of the diagnosis and provides easier access to the peripheral lung tissues, with the yield of about 75% with a lower pneumothorax rate than computed tomography (CT) guided fine needle aspiration [5, 13, 23]. However, currently there are no quantitative, established methods to evaluate EBUS images of peripheral pulmonary lesions. In this thesis, I will implement several approaches to quantify parameters that may improve pulmonary tissue characterization.

In this respect, image processing needs to be compared to a well-established technique. In our case, this technique will be histopathological analysis or the biopsy samples. EBUS-guided transbronchial needle aspiration (EBUS-TBNA) with a convex
probe allows for optimal sampling during image-guided biopsy procedures [8, 11]. Biopsies are performed with a needle and then analyzed by histology, which is currently the gold standard for tissue characterization. In Figure 2.1 the EBUS-TBNA apparatus is shown. The advantage of having ultrasound with the transbronchial needle is that the biopsy procedure can be done in real time.

Figure 2.1 Transbronchial needle with convex probe transducer. Retrieved from [24].

EBUS-guided FNA has provided statistically significant sensitivity and specificity rates for tissue diagnosis. For example, EBUS guided biopsy sampling had a sensitivity of 83% (range 45-100%), a specificity of 97% (range 88-100%) and a false-negative rate of 22% in a study that involved 1201 patients [6, 14]. The increased sensitivity and specificity allows for better classification of the tumor. The drawback of this method is that it is invasive and biopsy numbers should be reduced.
2.2 Grey scale Analysis

The previous studies have revealed that greyscale analysis can be useful in the diagnosis of malignant and benign tumors in lymphadenopathy [3, 4, 11, 12]. In greyscale analysis the EBUS images have pixel values ranging from 0-255, while humans can visualize 80 levels of grey [22].

*Figure 2.2 Human vision vs Computer vision of the grey scale*

The eye-vision image on the left of the Figure 2.2 demonstrates the levels of grey that the eye can see. The computer on the right demonstrates how the computer can see the full spectrum of grey. The advantage of being able to quantify the full spectrum will allow for more precise classification of the EBUS image. Previous studies based on raw EBUS intensity could not obtain sufficient classification power. The quantitative, statistical features from EBUS images have the potential to provide improved
characterization power. Figure 2.3 shows the summary of the quantitative analysis approaches with the overall goal of finding the best classifying method. First, the EBUS images are received from the clinic with the same settings of gain and contrast.

Next, basic first stage ROI’s are selected by the clinician from the images to be used as samples. Then automatized MATLAB algorithm selected each ROI. Next, statistical analysis was performed using histogram analysis, texture analysis, multiple regression and machine learning algorithms for each ROI. Below I detail the quantitative imaging parameters that I used to characterize the EBUS images.
2.3 Quantitative Parameters of Grey Scale Analysis

As for the grey scale analysis, I have applied histogram analysis and texture analysis for the characterization of EBUS images. Grey scale histogram analysis depends mainly on individual pixel values and not on their interaction with neighboring pixels [4, 11, 12, 17]. Histogram based parameters include: Mean, Median, Mode, Max, Min, Height, Width, Standard deviation, Kurtosis, Skewness and Height/Width ratio. The following parameters were applied by Morikawa et al. for the analysis of lung tissue: height (the number of maximum pixels), width (maximum-minimum gray value), height/width ratio [the number of maximum pixels/ (maximum-minimum gray value)], standard deviation of histogram, kurtosis and skewness [12]. They showed that tumor histogram shape was wide and rough with non-Gaussian shapes. In contrast for benign lesions, the histogram width was narrow, shape was sharp, and height was taller [12]. Thus, there were quantitative and qualitative features in the histogram distributions that could characterize the lesions.
**Histogram Analysis.**

Since we are interested in quantitative imaging parameters to characterize tissue, below I describe several histogram-related parameters.

*Figure 2.4 Pixel selection and Histogram analysis*

The histogram in Figure 2.4 shows the parameters of Mean, Median, Mode, Max, Min, Height, Width, Skewness and Kurtosis. Figure 2.4 left represents how the pixels of the image are treated individually as separate counts for the histogram analysis. The figure on the right demonstrates where the histogram shape and relevant histogram parameters that can be quantitated, which were also investigated by Morikawa *et al* [12]. The goal by quantifying these parameters is to find a feature that will separate the malignant image from the benign image by histogram analysis as a first step towards classification. During the histogram analysis, identical gain and contrast settings were assumed. Below I summarize the brief descriptions of the histogram parameters.
**Mean:** Mean (m) is a measure of the average intensity of an image [25]. The mathematical definition of mean is straightforward:

\[
m = \sum_{i=0}^{L-1} z_i p(z_i)
\]  

(2.1)

where, L=Number of possible intensity levels, 
\(z_i\)=random variable indicating intensity,  
p(z)=histogram of the intensity levels in a region.

**Standard deviation:** Standard deviation (\(\sigma\)) is a measure of average difference pixel to pixel [25].

\[
\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)}
\]  

(2.2)

where, L=Number of possible intensity levels,  
\(z_i\)=random variable indicating intensity p(z)=histogram of the intensity levels in a region,  

**Mode:** The most reoccurring pixel value in the image,  
**Median:** The middle pixel value in the image,  
**Max:** The largest pixel value in the image  
**Min:** The smallest pixel value in the image.

![Figure 2.5 Histogram features of Skewness and Kurtosis](image)
Skewness: Skewness ($\mu_3$) measures the sharpness in the peaks of the histogram compared to a normal Gaussian distribution.

$$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)$$

(2.3)

$\mu_3$=Skewness

$$\mu_4 = \sum_{i=0}^{L-1} (z_i - m)^4 p(z_i)$$

(2.4)

Kurtosis: Kurtosis ($\mu_4$) is a measure of how outlier prone a distribution is [26].

Range: Range is the Maximum pixel value subtracted by the minimum pixel value.

$$\lambda = (Pix_{max} - Pix_{min})$$

(2.5)

Texture Analysis

Texture analysis is based on co-occurrence matrices that have normalized sums of 1. The analysis is less affected by gain and contrast settings of the ultrasound images. The grey level co-occurrence matrix is formed by such matrix, which is derived from the relationship between neighboring pixels as well as the distance between them [11, 19]. The hypothesis is that malignant lesions are expected to have different texture characteristics than that of benign lesions.

Figure 2.6 How the Computer sees 5 different textural patterns [27]
The texture analysis example in Figure 2.6 shows how the different patterns are analyzed by the computer: The eye can easily discern the different patterns shown on the image on the left, the computer then a statistical representation of texture will be created for the benign and malignant EBUS images. A grey level co-occurrence matrix is one pixel and the eight pixels that surround it. The features that can be observed from grey level co-occurrence matrices are: Entropy, Homogeneity, Energy, Contrast and Correlation.

**Entropy**: Entropy is the statistical measure of disorder in the distribution of intensities or the randomness of the image. It is defined by the following equation:

\[
Entropy = -\sum_i \sum_j P_d(i,j) \log(P_d(i,j))
\]  

Where P is the grey-level co-occurrence matrix.

**Contrast**: Contrast is used to explain the difference in pixel intensities from one region of interest (ROI) of the image to another, and defined as:

\[
Contrast = \sum_i \sum_j (i - j)^2 P_d(i,j)
\]  

*Figure 2.7 Contrast example*
Figure 2.7 displays an example of contrast as the difference from one side of the image to another.

**Correlation:** Correlation is how neighboring pixel intensities are related to one another. It is a measure of how correlated a pixel is to neighbor over the entire image [25].

\[
Correlation = \frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y)P_d(i,j)}{\sigma_x \sigma_y}
\]  

(2.8)

Where \(\mu_x\) and \(\mu_y\) represent the mean of the rows and columns of an image

\(\sigma_x\) and \(\sigma_y\) represent the standard deviation of the rows and columns in an image.

**Energy:** Energy is the varieties of intensities found in the image. It is a measure of sum of squared elements in the image. Energy is the sum of all the different energies in the image.

\[
Energy = \sum_i \sum_j P_{d}^2(i,j)
\]

(2.9)

![Diagram of Low Energy vs High Energy](image)

*Figure 2.8 Low Energy vs High Energy*
Figure 2.8 above represents how the low energy and high energy are seen in an image. The low energy image has the energy levels that are the same or very few different energy levels. The high-energy image has many different energy levels.

**Homogeneity:** Homogeneity is the frequency at which near identical intensities are adjacent to each other [11]. Homogeneity describes how uniform the image appears. For example, if every pixel in the image had one intensity value or the same pattern of pixels was seen across the image then the image would be very homogeneous.

\[
Homogeneity = \sum_i \sum_j \frac{p_{ij}^2(\|i-j\|)}{1+|i-j|}
\]  

(2.10)

A low homogeneous image (shown in Figure 2.9) that has a variety of intensities and a high homogeneous image has intensities that are very close to each other in value. Homogeneity is inversely correlated with energy as high homogeneity is also low energy and vice versa.
2.4 Receiver Operating Curve Analysis

The ROC curve is a measure of how well a parameter or feature classifies. The ROC curve is obtained by plotting the true positive rate (Sensitivity) vs the false positive rate (1-Specificity). A curve that passes through the point (0, 1) would indicate the perfect classification. A test that would have a very low classifier would follow the diagonal line on the figure 3.9 [28–36].

![Figure 2.10 Different types of ROC curves and how they assess the model [28]](image)

The ROC curves shown in Figure 2.10 represents the classification ability of a feature. The ROC curve represents the accuracy of a model in classifying between malignant and benign lesions.
The AUC is an important parameter in assessing the accuracy of the model that is used for classification. The AUC is obtained from the ROC curve, representing the summed area. For example, if the area is between 0.9 and 1.0, then the classification rank is “Excellent” (Figure 2.11).

<table>
<thead>
<tr>
<th>AUC</th>
<th>Classifier Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9-1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>Good</td>
</tr>
<tr>
<td>0.7-0.8</td>
<td>Fair</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>Poor</td>
</tr>
<tr>
<td>0.5-0.6</td>
<td>Bad</td>
</tr>
</tbody>
</table>

*Figure 2.11 The Rankings of the classifiers based on the Area Under the Curve*

The AUC is an important parameter in assessing the accuracy of the model that is used for classification. The AUC is obtained from the ROC curve, representing the summed area. For example, if the area is between 0.9 and 1.0, then the classification rank is “Excellent” (Figure 2.11).

*Figure 2.12 Cutoff points [28, 35]*

Figure 2.12 shows True Positive, False Positive, True Negative and False Negative with chosen cutoff point. [30, 37]. From the cutoff point, an analysis table called a confusion matrix can be formed, as seen below [36].
Table 2.1 Confusion Matrix

<table>
<thead>
<tr>
<th>Test</th>
<th>Present</th>
<th>n</th>
<th>Absent</th>
<th>n</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>True Positive(TP)</td>
<td>a</td>
<td>False Positive (FP)</td>
<td>c</td>
<td>a+c</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative(FN)</td>
<td>b</td>
<td>True Negative(TN)</td>
<td>d</td>
<td>b+d</td>
</tr>
<tr>
<td>Total</td>
<td>a+b</td>
<td></td>
<td>c+d</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Table 2.1 the several other statistically relevant parameters can be derived [30, 37]. The first one is the sensitivity, which is the probability that the test will be positive when the disease is present [33, 37]. Sensitivity is used to assess how well the model predicts malignancy when the EBUS image is actually malignant.

\[
\text{Sensitivity} = \frac{a}{a+b} \quad (3.11)
\]

Next, the specificity, which is the probability that the test will be negative when the disease is not present [33, 37]. Specificity is used to assess how well the model predicts benign tissue when the EBUS image is actually benign.

\[
\text{Specificity} = \frac{d}{c+d} \quad (3.12)
\]

Positive Likelihood ratio (+LR): ratio between the probability of a positive test given the present of disease and the probability of a positive test given the absence of the disease [33, 37].

\[
+LR = \frac{\text{Sensitivity}}{1-\text{Specificity}} \quad (3.13)
\]

\[
-\text{LR} = \frac{1-\text{Specificity}}{\text{Sensitivity}} \quad (3.14)
\]
Negative Likelihood ratio (-LR): ratio between the probability of a negative result given the presence of the disease and the probability of a negative test result given the absence of the disease [37].

The Youden Index is a measure of performance of a classifying test. The test ranges from -1 to 1 with 0 meaning that there is a 50-50 chance of getting the right classification. In other words, if there is a value of 0 or less the test is useless and if the value is 1 there is perfect classification [38].

\[ J = \text{Sensitivity} + \text{Specificity} - 1 \]  

(3.15)

Accuracy is how close the classification is to the true value of the data class. Accuracy is a useful parameter for identifying what the best sensitivity and specificity is needed for the classification method [34].

\[ \text{Accuracy} = \frac{a+d}{a+b+c+d} \]  

(3.16)

2.5 Multiple Regression

Multiple regression involves combining multiple features to achieve better sensitivity and specificity for improved classification power (i.e. higher sensitivity and specificity parameters). In this thesis, I will use the parameters obtained by the first, second order image features and histogram analysis. One should be careful in combining parameters to perform multiple regression, because some parameters can be highly correlated that can result in overfitting. To determine whether parameters are correlated, one can simply perform a correlation analysis and obtain correlation matrix [18, 36, 39, 40].
Table 2.2 Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Contrast</th>
<th>Energy</th>
<th>Homogeneity</th>
<th>Height</th>
<th>H/W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>1</td>
<td>-0.3337</td>
<td>0.4173</td>
<td>0.4569</td>
<td>-0.3647</td>
<td>-0.3784</td>
</tr>
<tr>
<td>Contrast</td>
<td>-0.3337</td>
<td>1</td>
<td>-0.7519</td>
<td>-0.8797</td>
<td>0.9709</td>
<td>0.9059</td>
</tr>
<tr>
<td>Energy</td>
<td>0.4173</td>
<td>-0.7517</td>
<td>1</td>
<td>0.9482</td>
<td>-0.7344</td>
<td>-0.694</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.4569</td>
<td>-0.8797</td>
<td>0.9482</td>
<td>1</td>
<td>-0.8568</td>
<td>-0.8054</td>
</tr>
<tr>
<td>Height</td>
<td>-0.3647</td>
<td>0.9709</td>
<td>-0.7344</td>
<td>-0.8568</td>
<td>1</td>
<td>0.973</td>
</tr>
<tr>
<td>H/W</td>
<td>-0.3784</td>
<td>0.9059</td>
<td>-0.694</td>
<td>-0.8054</td>
<td>0.973</td>
<td>1</td>
</tr>
</tbody>
</table>

The correlation matrix displayed in Table 2.2 shows the features that cannot be combined due to high correlation. From my analysis, I obtained the features that cannot be combined for multiple regression are Homogeneity and Energy, Height and Contrast, Height/Width and Contrast, and Height/Width and Height. The combination of features with high correlation leads to overfitting the data because it is essentially the same as adding the same feature twice to the data.

The equation for multiple regression is as follows:

\[ Y = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_n X_n + \epsilon \]  \hspace{1cm} (3.17)

Where \( Y \) = Classifier, \( B_0 \) = intercept, \( X_0 \) and \( X_1 \) = the features to be combined = the standard error of the method.
Multiple regression as seen in Figure 2.13 is the process of projecting points on to a plane of best fit [41]. After the points are projected onto a plane cutoff points can be determined for classification of the data.

2.6 Machine Learning

<table>
<thead>
<tr>
<th>AUC</th>
<th>Classifier Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9-1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>Good</td>
</tr>
<tr>
<td>0.7-0.8</td>
<td>Fair</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>Poor</td>
</tr>
<tr>
<td>0.5-0.6</td>
<td>Bad</td>
</tr>
</tbody>
</table>

Machine learning is the creation of an algorithm that takes past data to complete either classification or regression on future data in the form of prediction. Machine learning is
on the frontier of statistical analysis and it requires the organization of the initial data into
feature vectors to be analyzed. As an advanced approach, machine learning is also applied
to check whether it will perform better than Multiple Regression, Histogram and Texture
Analysis as shown in Figure 2.14. The goal of machine learning is to create a machine
that learns how to classify data accurately [11, 42]. Machine Learning uses hyperplanes
(lines), decision trees and clustering algorithms to separate raw data into classes. Machine
learning might be better suited for the cases, where one tries to find the difference
between two features to classify malignant and benign images. The same chart that was
shown in Figure 2.11 is also shown in Figure 2.14 to indicate our aim that we want to
increase AUC values with the help of machine learning algorithms.

Machine learning has the ability to take many features (imaging parameters) and use
them to classify each image. An example where machine learning can be very useful is
when one wants to assess the failure rate of a machine. For example, the question that
needs to be solved could be how long will it take for the machine to fail? The features
can include: Average fail rate of that machine, age of the machine, maintenance and
history of malfunctions of that machine. Notice that all the features are relevant to the fail
rate of the machine, however one feature alone will not describe the length of time that it
will take for that machine to fail. The use machine learning can assess all the features
together and give an accurate prediction of when that machine will be expected to fail.
The two outputs of this process would be the machine is still working and the machine
failed.

Another way that machine learning can be used is to identify the features that cause
a particular outcome. In the case of this thesis, this is the method that is being used
because the desired outcome is to find the best features for determining the malignancy or the benignity of the EBUS images. The goal of selecting features is to find the features that best represent the classifications, then taking those features and combining them for machine learning.

The different types of sample selection for the training data set include **cross validation**, the **holdout method** and the **no validation method**. The **cross-validation method** consists of selecting a set number of times that the data will be folded to test the last sample in the set. This method is a recursive method that takes the initial samples as training to test the last sample of the data [42–44]. Then the second to last sample is tested and so on. This method fits the data many times and would not be appropriate for large data sets. The hold out method involves holding out a percentage of random samples from the dataset for training, this method is best for large data sets due to the removal of data in the test set. The model can then be tested against the remaining data to assess accuracy. The hold out method is the ideal method for creating a machine that can be tested on the true population. The cross-validation method would not be good for testing on the true population because it fits a small sample set many times. The no-validation method uses the entire dataset for training and testing. This method does not suit our needs due to the high artificially high yield it will produce to have the same dataset in the model and the training set. In this thesis, the cross-validation method was chosen to analyze the data.

Machine learning has different categories of learning, in this thesis I used “supervised learning” scheme. Supervised learning is where the input dataset has a specific structure of inputs and desired outputs, which can be used for medical data analysis in the event of benign and malignancy. Supervised learning is used to train a
model to discriminate between different classes of variables using features that are innate to each class. [42].

There are many machine learning methods out there that can be used to classify data, however in this thesis I only used Support Vector Machines. Support vector machines were used in this thesis because they produced the best classification results for the data set that I had.

2.6.1 Support vector machine (SVM) Method

This method employs the use of a hyper plane to separate the data into two or more classes. The goal of the SVM is to select the most optimal hyper plane that separates the data the best. Accuracy is calculated by identifying the amount of data that was correctly classified by the optimal hyper plane.

Linear SVM

The linear model employs the use of a linear hyper plane, which separates the data into two or more classifications. This is illustrated by Figure 2.15.
Figure 2.15 Linear SVM Example

Figure 2.15 shows the linear SVM separates the data with a linear line. The red and blue points represent different classes of data and the $X_1$ and $X_2$ represent 2 features that are being observed [42].

**Quadratic SVM**

The quadratic method uses a quadratic line to separate the data into the two specified classes. The only difference from the linear SVM is that a quadratic line is used to separate the data.

**Cubic SVM**

The cubic method employs the use of a cubic line or kernel to separate the data.

**Gaussian SVM**

The Gaussian method applies a Gaussian line or kernel to separate the data.
Radial SVM

The radial method uses a closed circle to separate the data into 2 classes. This method is illustrated shown in Figure 2.16.

![Figure 2.16 Cubic SVM and Radial SVM](image)

Figure 2.16 shows the cubic SVM (on right) separates the data with cubic lines. The radial SVM (on left) separates data with a boundary. The red and blue points represent different classes of data and the $X_1$ and $X_2$ represent 2 features that are being observed [42].

Summary SVM

SVMs use a hyper plane to separate and classify data. There are an infinite number of hyper planes that can distinguish between the points on a graph however there is only one optimal hyper plane that gives the most accurate results [42]. The SVM aims to output the hyper plane with the most optimal results. Then the confusion matrix (which gives sensitivity and specificity) and the Receiver Operating Curve are used to interpret the results of the hyper plane. The cubic and Gaussian methods were used in this thesis because they provided the most accurate results.
2.7 EBUS ROI selection

The EBUS images were received from Kassem Harris MD. The EBUS images all had the same settings of gain of 18/19 and contrast of 5/8. The same settings are critical in image analysis for accurate comparison between the images. Approximate region of interests (ROIs) were initially determined by Kassem Harris MD. The initial ROI provided borders for my automatized, MATLAB based ROI selection for each lesion, on a Dell OptiPlex 9020 desktop with a Core i7 vPro running Windows 7 Professional. ROI analysis was completed in two methods: one method with the whole lesion as an ROI and a second method with 10 ROI’s selected in the tumor.

2.7.1 Whole lesion ROI

In our main ROI selection approach, I took the whole lesion as an ROI and extract the features accordingly from that particular ROI. There were 13 total EBUS images from 13 patient data for the analysis. A MATLAB program was automatized to select the ROI’s for the most representative areas of the lesions, initially defined by Dr Harris. Next, a feature vector representing statistical image features were calculated from each ROI. The goal of the statistical analysis was to identify the cut point to separate the ROI’s as either malignant or benign tumors.

2.7.2 Multiple ROI

Since the sample size (N=13) was small for testing the machine learning algorithms, it was decided to obtain 10 additional ROIs arbitrarily from each sample, which lead to N of 130 artificially generated ROIs. The sample size number kept constant to represent each lesion with the same number of weight. The sizes of the ROIs were variable because
each lesion size was different. In order to fit 10 ROIs in the lesion of each image the size of the ROI must change to fit the whole lesion. I have applied the methods described in the previous chapter to obtain these results. Briefly, ROIs were evaluated in MATLAB (2015a) for histogram and texture grey scale analysis. The features that were extracted using histogram grey scale analysis were: mean, median, mode, max, min, max-min, and standard deviation pixel to pixel. The features that were extracted using texture grey scale analysis were: entropy, correlation, contrast, energy, homogeneity. The extracted features using histogram analysis were: height, width, kurtosis, skewness, and height-width ratio. The features (quantitative parameters) were exported from MATLAB to a text file then stored into a feature vector in Microsoft Excel. The feature vector was then imported into MedCalc statistical software analysis package for ROC analysis. MedCalc provides the area under the curve (AUC), sensitivity, specificity, Youden index, +LR, -LR and accuracy of each feature.

2.8 Machine Learning setup

2.8.1 Constructing the Feature Vector

The layout of a feature vector consists of the first column in binary, in the form of 1s and 0s, where benign is 1 and malignant is 0, while the rest of columns are the predictors, where the predictors are the texture and histogram features. The rows of the feature vector represent each ROI that was extracted in the grey scale analysis. The machine is then used to calculate the accuracy of the data and output to a confusion matrix, a ROC curve and scatter plot to visualize the data. The MATLAB toolbox has also the visualization
feature, which was used to select the best kernel for the support vector machine (SVM) to use to classify the data.

2.8.2 Cross Validation and Hold-Out Methods

The classificationLearner © GUI in MATLAB allows the user to analyze data via cross validation or hold out methods. Cross validation is the optimal choice for small data sets as it allows the folding of data and gives a more accurate training machine for small data sets. My data consists of 13 samples representing the whole lesion dataset and a 130-samples obtained by using 10 ROIs. The cross-validation method was performed on both 13 and 130 sample data sets to assess the accuracy of the classification.

The hold out method involves selecting a percentage of the data; in this thesis 25% of the data used for training to be tested against the remaining data (75%). This method is useful for large sample sizes, but would not be appropriate for small samples sizes such as N=13 samples.

2.8.3 Machine Selection

As I indicated previously in the Methods section, the machine selection is important for the classification of the data, and MATLABs classification Learner © is that a GUI gives the user the ability to choose between decision trees, support vector machines, k nearest neighbors and ensemble classifiers. Decision trees are similar to “if-then” statements that can be created by a script file. SVM was chosen due to the use of a hyper-plane to separate the data gave the best classification accuracy over the other machine learning methods. The classification Learner allows for quick access to the different types kernels that can
be used for the SVM. After testing each type of SVM I found that the cubic SVM and the Gaussian SVM produced the most accurate results compared to the other types of SVMs.

2.8.4 Training and exporting the Machine

Training the SVM can be done by importing a feature vector into classificationLearner®, then a predictor variable must be selected and the type of sampling method must be selected. The sampling methods are cross-validation and hold-out methods. The model is then trained according to that method. In the hold-out method 25% of the data is used to train the model and in the cross-validation method the entire dataset is used to train and test on. The cross-validation method was used to train the machines because it produced the best results.

Once the machine is trained on the data the model can be exported into the MATLAB workspace and it can be used to predict other EBUS images. The prediction completed by taking the trained model and the new feature vector of the image to be classified and creating a new predicted response vector of the EBUS image to be classified. The predicted vector is then compared to the actual values and a confusion matrix is created as shown in Table 2.1 detailed in the previous chapter.
3 RESULTS

3.1 ROI Selection of the EBUS Images

As a first approach, MATLAB is used and an automatized single ROI representing the whole lesions was obtained, which ensured that the entire lesion was counted in the analysis. Figure 3.1 shows a representative malignant EBUS image in Fig 3.1a. Then the MATLAB generated ROI is placed containing the whole lesion (Fig 3.1b). After the data is collected from the ROI a histogram is generated (Fig 3.1c) that indicates the gray level intensities in the image.

(a) Lymph node: adenocarcinoma. (b) Malignant Image with whole lesion ROI selected.

(c) Histogram of the ROI.

Figure 3.1 Malignant EBUS image whole lesion method
The representative histogram of the malignant EBUS showed a sharp peak and a height level was 825 counts. The height in a histogram is subject to change depending on the size of the tumor (number of pixels in the image) in the EBUS image.

(a) Original Image: Lymph node:   (b) Benign image with whole lesion ROI.  

(c) Histogram of ROI.

*Figure 3.2 Benign EBUS image whole lesion method*

The representative histogram of the whole ROI from of EBUS image from a benign lesion is shown in Figure 3.2(c). It has about the same width as the malignant histogram in Figure 3.1(c). The heights of these two histograms however are different by about 500 counts. This difference is due to the size of the malignant tumor as it is larger than that
of the benign tumor. Although it is useful to quantify histogram counts, as indicated above, this differences may not be true classifier for tissue characterization due to differences in image sizes. Thus, the number of pixels in each ROI should be kept constant for the analysis.

Figure 3.3 Malignant EBUS image multiple ROI method (a) Original Image: Lymph node: adenocarcinoma (b) Image from malignant lesion with 10 ROIs selected. (c, d) Histograms
Next, the malignant EBUS image was divided into multiple ROI’s to increase the number of samples taken in the image (Figure 3.3). The representative histograms (only 8 is shown below) have sharp peaks.

(a) Original Image: Benign lymph node. (b) Benign image with 10 ROI selected.

(c) Histograms of ROI (1-4) (d) Histograms of ROI (5-8)

Figure 3.4 Benign image multiple ROI method

A representative benign image was divided into multiple ROI’s as shown in Figure 3.4, which indicates the height parameter distributions were lower than that of malignant case.
3.2 MedCalc Analysis of the Image Parameters (N=13)

*Table 3.1 13 sample MedCalc Results*

<table>
<thead>
<tr>
<th>Features</th>
<th>Cut point</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>Positive LR</th>
<th>Negative LR</th>
<th>AUC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>≤58.82</td>
<td>57.14</td>
<td>100</td>
<td>0.5714</td>
<td>-</td>
<td>0.43</td>
<td>0.786</td>
<td>0.0532</td>
</tr>
<tr>
<td>Median</td>
<td>≤58.7</td>
<td>57.14</td>
<td>100</td>
<td>0.5714</td>
<td>-</td>
<td>0.43</td>
<td>0.786</td>
<td>0.0532</td>
</tr>
<tr>
<td>Mode</td>
<td>≤51.2</td>
<td>57.14</td>
<td>100</td>
<td>0.5714</td>
<td>-</td>
<td>0.43</td>
<td>0.762</td>
<td>0.0772</td>
</tr>
<tr>
<td>Max</td>
<td>≤104.2</td>
<td>71.43</td>
<td>100</td>
<td>0.7143</td>
<td>-</td>
<td>0.29</td>
<td>0.786</td>
<td>0.0603</td>
</tr>
<tr>
<td>Min</td>
<td>≤31.8</td>
<td>85.71</td>
<td>83.33</td>
<td>0.6905</td>
<td>5.14</td>
<td>0.17</td>
<td>0.786</td>
<td>0.0628</td>
</tr>
<tr>
<td>Contrast</td>
<td>≤308.65</td>
<td>85.7</td>
<td>83.33</td>
<td>0.6905</td>
<td>5.14</td>
<td>0.17</td>
<td>0.738</td>
<td>0.1791</td>
</tr>
<tr>
<td>Correlation</td>
<td>≥-0.0059</td>
<td>42.9</td>
<td>83.33</td>
<td>0.2619</td>
<td>2.57</td>
<td>0.69</td>
<td>0.548</td>
<td>0.7868</td>
</tr>
<tr>
<td>Height</td>
<td>≥41.9</td>
<td>100</td>
<td>66.7</td>
<td>0.6667</td>
<td>3</td>
<td>0</td>
<td>0.786</td>
<td>0.0867</td>
</tr>
<tr>
<td>Width</td>
<td>≤97.1</td>
<td>57.1</td>
<td>83.3</td>
<td>0.4048</td>
<td>3.43</td>
<td>0.51</td>
<td>0.595</td>
<td>0.5977</td>
</tr>
<tr>
<td>H/W</td>
<td>≥0.5237</td>
<td>100</td>
<td>83.3</td>
<td>0.8333</td>
<td>6</td>
<td>0</td>
<td>0.857</td>
<td>0.0137</td>
</tr>
<tr>
<td>STD</td>
<td>≤3.211</td>
<td>100</td>
<td>50</td>
<td>0.5</td>
<td>2</td>
<td>-</td>
<td>0.667</td>
<td>0.3624</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>≥3.77</td>
<td>42.86</td>
<td>83.33</td>
<td>0.2619</td>
<td>2.57</td>
<td>0.69</td>
<td>0.524</td>
<td>0.8922</td>
</tr>
<tr>
<td>Skewness</td>
<td>≤0.344</td>
<td>28.57</td>
<td>100</td>
<td>0.2857</td>
<td>-</td>
<td>0.71</td>
<td>0.548</td>
<td>0.7913</td>
</tr>
<tr>
<td>Max-min</td>
<td>≥76.6</td>
<td>71.43</td>
<td>50</td>
<td>0.2143</td>
<td>1.43</td>
<td>0.57</td>
<td>0.548</td>
<td>0.7917</td>
</tr>
<tr>
<td>Entropy</td>
<td>≥5.9186</td>
<td>28.6</td>
<td>100</td>
<td>0.2857</td>
<td>-</td>
<td>0.71</td>
<td>0.571</td>
<td>0.6829</td>
</tr>
<tr>
<td>Energy</td>
<td>≤0.0005</td>
<td>85.7</td>
<td>83.3</td>
<td>0.6905</td>
<td>5.14</td>
<td>0.17</td>
<td>0.762</td>
<td>0.125</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>≤0.1319</td>
<td>85.7</td>
<td>83.3</td>
<td>0.6905</td>
<td>5.14</td>
<td>0.17</td>
<td>0.738</td>
<td>0.1791</td>
</tr>
</tbody>
</table>
The best classifiers in the data set that involves selecting whole region as ROI (N=13) were Homogeneity, Energy, Contrast, Height, Height-Width, Mean, Median and Minimum pixel value as shown bold in Table 3.1 I have taken only Homogeneity, Energy, Contrast and Min as the best classifiers since the other parameters (Height, Height-Width ratio, Mean and Median) are highly correlated. For example, if the tumor of the malignant EBUS image is large then the height of the histogram will be greater than the height of the benign EBUS histogram. This creates a problem because the size of the tumor was mostly larger in the malignant lesions than the benign ones. The ROC curves of uncorrelated features are shown in figure 4.5. The L-shaped plots and high AUC numbers (~0.74, 0.76, 0.74, 0.79, respectively) indicate that these features (imaging parameters) are “fair” classifiers (see Table 3.1).
Each one of the curves displayed in Figure 3.5 have an area under the curve that falls within the **fair** category that is shown in Figure 2.14. Since the whole ROI-based selection provided only N=13 samples, the low classification power (low AUC) might be due to this low tissue sample number. We need to obtain higher AUC values before these parameters could be considered as good classifiers.
3.3 MedCalc Analysis of the Image Parameters (N=130)

The best classifiers in the data set that involves selecting sub-ROIs (N=130) were
Homogeneity, Energy, Contrast, Min, Height and Height/Width ratio as shown in bold in Table 3.2.

*Table 3.2 130 sample MedCalc Results*

<table>
<thead>
<tr>
<th>N=130</th>
<th>Features</th>
<th>Cut point</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>+ LR</th>
<th>- LR</th>
<th>AUC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>≤64.72</td>
<td>64.3</td>
<td>71.7</td>
<td>0.3595</td>
<td>2.27</td>
<td>0.5</td>
<td>0.656</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>≤63</td>
<td>58.6</td>
<td>70</td>
<td>0.2857</td>
<td>1.95</td>
<td>0.59</td>
<td>0.647</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td>Mode</td>
<td>≤59</td>
<td>67.1</td>
<td>60</td>
<td>0.2714</td>
<td>1.68</td>
<td>0.55</td>
<td>0.657</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>≤119</td>
<td>74.3</td>
<td>50</td>
<td>0.2429</td>
<td>1.49</td>
<td>0.51</td>
<td>0.62</td>
<td>0.0149</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>≤38</td>
<td>84.3</td>
<td>58.3</td>
<td>0.4262</td>
<td>2.02</td>
<td>0.27</td>
<td>0.721</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>≥334.224</td>
<td>81.4</td>
<td>81.7</td>
<td>0.631</td>
<td>4.44</td>
<td>0.23</td>
<td>0.752</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>≥-0.036</td>
<td>74.3</td>
<td>36.7</td>
<td>0.1095</td>
<td>1.06</td>
<td>0.86</td>
<td>0.515</td>
<td>0.7791</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>≥0.4767</td>
<td>92.9</td>
<td>63.3</td>
<td>0.5619</td>
<td>2.53</td>
<td>0.11</td>
<td>78.3</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Width</td>
<td>≤100</td>
<td>55.7</td>
<td>68.3</td>
<td>0.2405</td>
<td>1.76</td>
<td>0.65</td>
<td>0.597</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>H/W</td>
<td>≥45</td>
<td>97.1</td>
<td>58.3</td>
<td>0.5548</td>
<td>2.33</td>
<td>0.049</td>
<td>0.787</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>≤4.205</td>
<td>94.3</td>
<td>16.7</td>
<td>0.1095</td>
<td>1.13</td>
<td>0.34</td>
<td>0.54</td>
<td>0.4393</td>
<td></td>
</tr>
<tr>
<td>Kurtosis</td>
<td>≥2.594</td>
<td>71.4</td>
<td>43.3</td>
<td>0.1476</td>
<td>1.26</td>
<td>0.66</td>
<td>0.515</td>
<td>0.766</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>≤-0.0486</td>
<td>54.3</td>
<td>61.7</td>
<td>0.1595</td>
<td>1.42</td>
<td>0.74</td>
<td>0.529</td>
<td>0.5766</td>
<td></td>
</tr>
<tr>
<td>Max-min</td>
<td>≥60</td>
<td>90</td>
<td>26.7</td>
<td>0.1667</td>
<td>1.23</td>
<td>0.38</td>
<td>0.562</td>
<td>0.2317</td>
<td></td>
</tr>
<tr>
<td>Entropy</td>
<td>≥5.793</td>
<td>50</td>
<td>66.7</td>
<td>0.1667</td>
<td>1.5</td>
<td>0.75</td>
<td>0.575</td>
<td>0.1382</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>≤0.0006</td>
<td>85.7</td>
<td>83.3</td>
<td>0.6905</td>
<td>5.14</td>
<td>0.17</td>
<td>0.754</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Homogeneity</td>
<td>≤0.1383</td>
<td>84.3</td>
<td>71.7</td>
<td>0.5595</td>
<td>2.97</td>
<td>0.22</td>
<td>0.747</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>
The ROC curves of these features are shown in Figure 3.6. For the ROC analysis, height, height-width ratio, mean and median were not included due to high correlation between the parameters.

![ROC Curves for Homogeneity, Energy, Contrast, and Minimum](image)

*Figure 3.6 ROC Homogeneity, Energy, Contrast and Minimum multiple ROI*

Again for this data set, the AUC curves fall within the *fair* category. To check the improvement in the classification, I tested the multiple regression method that involves combining several parameters for the classification, as results are detailed below.
3.4 Multiple Regression Method (N=130-samples)

N=130 sampled data was used for the multiple regression, as N=13 was too small of a sample size to test this method. If individual parameter represents a single dimension, the multiple regression allows combining multiple variables (dimensions) or features to a lower dimension for better classification.

MedCalc analysis was used for the multiple regression analysis. Features that had a high correlation coefficient were not combined to eliminate overfitting possibilities. The features that had the lowest p-value, highest area under the curve, specificity and sensitivity were combined for multiple regression analysis. The features combined in multiple regression were:

- 1) Contrast and Energy
- 2) Min, Contrast and Energy
- 3) Min, Contrast and Homogeneity

Energy and Homogeneity were not combined because their correlation coefficient was high (r=0.95), as shown in table 3.2. Multiple regression in MedCalc produced predicted values summarized in Table 4.3 below. ROC curves are plotted below in Figure 3.7.
Figure 3.7 The ROC curves of the multiple regression combinations

The combinations in the multiple regression (Contrast and Energy; Min, Contrast and Energy; Min Contrast and Homogeneity) had the ROC curves, as shown in Figure 3.7, and AUC values of 0.86, 0.89 and 0.86, all fall under the good classifier category. The results show that the multiple regression analysis increased the classification accuracy.
Table 3.3 Combined features with Cut point, Sensitivity, Specificity and Youden Index

<table>
<thead>
<tr>
<th>Features</th>
<th>Cut point</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast Energy</td>
<td>≥0.6568</td>
<td>70</td>
<td>100</td>
<td>0.7</td>
</tr>
<tr>
<td>Min Contrast Energy</td>
<td>≥0.6392</td>
<td>71.43</td>
<td>95</td>
<td>0.6643</td>
</tr>
<tr>
<td>Min Contrast Homogeneity</td>
<td>≥0.5897</td>
<td>71.43</td>
<td>88.33</td>
<td>0.5976</td>
</tr>
</tbody>
</table>

Table 3.4 Combined features with LR, AUC and p-value

<table>
<thead>
<tr>
<th>Features</th>
<th>Positive LR</th>
<th>Negative LR</th>
<th>AUC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast Energy</td>
<td>-</td>
<td>0.3</td>
<td>0.86</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Min Contrast Energy</td>
<td>14.29</td>
<td>0.3</td>
<td>0.887</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>Min Contrast Homogeneity</td>
<td>6.12</td>
<td>0.32</td>
<td>0.858</td>
<td>&gt;0.0001</td>
</tr>
</tbody>
</table>

The Multiple Regression indicates that all of these combined features are “good” classifiers. The multiple regression method with the combination of Minimum pixel value, Contrast and Energy is the best classifier as it has the best area under the curve. The three combinations of parameters in the table above are all good classifiers. In the next section we want to test whether we can increase the classification power by using a Machine Learning approach.
3.5 MATLAB Machine Learning on the ROI datasets

Figure 3.8 Energy vs Contrast scatter plot

Figure 3.8 demonstrates an example of the features of Energy vs Contrast in a scatter plot for the 130 samples obtained by multiple ROIs. The scatter plot allows for visualization and identification of a specific support vector machine (SVM) that can be used to classify the data. The blue and red points represent the benign and malignant lesions, respectively. Among SVM methods, the cubic SVM showed the best accuracy for this scatter plot. Since I could not use multiple regression for the N=13 sample, I tested machine learning for this sample with cross-validation method. Machine learning allows us to view the combination of parameter listed in section 3.4 (e.g. Energy-Contrast, etc) for the N=13 sample method by using cross-validation, as detailed below.
Machine Learning with N=13 samples

**Figure 3.9 SVM ROC of the Min Contrast and Energy full lesion method**

Figure 3.9 displays the ROC curve for the full lesion data set (N=13) with an AUC of 0.98 which states that Min, Contrast and Energy, falls in the “excellent” category for classifying between malignant and benign EBUS images in this dataset.

**Figure 3.10 Confusion Matrix of the Min Contrast and Energy full lesion method**

<table>
<thead>
<tr>
<th>True Class</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>85.7%</td>
</tr>
<tr>
<td>0.0%</td>
<td>14.3%</td>
</tr>
</tbody>
</table>
The confusion matrix associated with the features of Min, Contrast and Energy is shown in the Figure 3.10. The sensitivity is found to be 100% and specificity is ~86% in this confusion matrix. Where the benign is represented as 0’s and the malignant is represented as 1’s. Although this is a promising result, it uses cross validation method, which is inherently exhibit overfitting by validating by its own data subset.

Next, I tested the combination of features of Min, Contrast and Homogeneity.

Figure 3.11 SVM ROC of the Min Contrast and Homogeneity full lesion method

The perfect “upside down L” like shape of the ROC curve in Figure 3.11 indicates that AUC of this combination is 1. This implies that all the malignant tumors have been
correctly identified. The results and accuracy of this method fall within the excellent classifier ranking. This is also shown in the confusion matrix below in Figure 3.12.

![Confusion Matrix Min Contrast Homogeneity 13 sample](image)

*Figure 3.12 Confusion Matrix Min Contrast and Homogeneity full lesion method*

Although we have perfect ROC curve, the confusion matrix output from the machine learning algorithm indicates, as in Figure 3.12, that this combination has a type 1 error, meaning that there is one sample that is **predicted malignant** but is **truly benign**.

**Machine Learning with N=130 samples**

This method employed the use of the N=130 ROI data set. Cross validation was used to assess the accuracy of the training model in MATLAB’s classification learner. The confusion matrix and ROC curve of Contrast and Energy are shown in Figure 3.13 and Figure 3.14, respectively.
The Machine learning method with Contrast and Energy combined produced an AUC of 0.96, which falls within the excellent category for classification as shown in Figure 3.13.
The confusion matrix and ROC curve for Contrast, Min and Energy is shown in Figure 3.15 and Figure 3.16, where the cubic SVM produced an AUC of 0.95, which again falls in the **excellent** category.
The confusion matrix and ROC curve of Min, Contrast and Homogeneity (shown in Figure 3.17 and Figure 3.18, respectively) produced an AUC of ~0.90. The results of the Machine Learning SVM methods are shown in Table 3.5.
Table 3.5 SVM Results

<table>
<thead>
<tr>
<th>ML Features</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min Contrast Energy (Gauss) n=13</td>
<td>92.3%</td>
<td>100</td>
<td>85.7</td>
<td>0.857</td>
<td>0.976</td>
</tr>
<tr>
<td>Min Contrast Homogeneity (Gauss) n=13</td>
<td>92.3%</td>
<td>83.3</td>
<td>100</td>
<td>0.833</td>
<td>1.00</td>
</tr>
<tr>
<td>Contrast Energy (Cubic) n=130</td>
<td>90.00%</td>
<td>96.7</td>
<td>84.3</td>
<td>0.81</td>
<td>0.961</td>
</tr>
<tr>
<td>Min Contrast Energy (Cubic) n=130</td>
<td>86.20%</td>
<td>78.3</td>
<td>92.9</td>
<td>0.912</td>
<td>0.951</td>
</tr>
<tr>
<td>Min Contrast Homogeneity (Cubic) n=130</td>
<td>81.50%</td>
<td>75</td>
<td>87.1</td>
<td>0.621</td>
<td>90.1</td>
</tr>
</tbody>
</table>

For the sample size of N=13, the combination of Min, Contrast and Energy and the combination of Min, Contrast and Homogeneity had an accuracy of 92.3%, sensitivities of 100% and 83.3%, respectively, and the specificities of 85.7% and 100%, respectively (summarized in Table 3.5). The areas under the curves were 0.98 and 1, respectively which indicate that these features were highly successful for classifying lesions.

For the sample size of N=130, the combination of Energy and Contrast was found to be the optimal for the classification, as compared to single feature (parameter) analysis. As Table 3.5 indicates, the accuracy of this method was 90% with a sensitivity and specificity of 96.7 and 84.3, respectively. The AUC was 0.961, which is very close to 1, indicating that these features are good classifiers. The second set of features used with this sample was Min, Contrast and Energy. These features gave the best classification of the images, as indicated by the Youden index of 0.912, which is an indicator of how well the sensitivity and specificity reflect the positive classification of the data (Malignant).
and the negative classification of the data (Benign). The area under the curve was 0.95, which is also close to 1 indicating a very good classification. The last set of features investigated were Min, Contrast and Homogeneity, which showed an accuracy of 81.5, sensitivity and specificity of 75 and 87.1, respectively. These features did not give as good of a classification as the other features discussed earlier, however these features still fall in the **excellent** category as a classifier with a Youden index of 0.65 and the AUC of 90.1.

<table>
<thead>
<tr>
<th>AUC</th>
<th>Classifier Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9-1.0</td>
<td>Excellent</td>
</tr>
<tr>
<td>0.8-0.9</td>
<td>Good</td>
</tr>
<tr>
<td>0.7-0.8</td>
<td>Fair</td>
</tr>
<tr>
<td>0.6-0.7</td>
<td>Poor</td>
</tr>
<tr>
<td>0.5-0.6</td>
<td>Bad</td>
</tr>
</tbody>
</table>

*Figure 3.19 Machine Learning, Multiple Regression and Histogram/Texture Rankings*

Figure 3.19 represents the improved progression of the statistical analysis from Histogram and Texture Analysis to Multiple Regression to Machine Learning. The objective of the machine learning method was not only to increase the AUC but also to increase the sensitivity and specificity of the classification. The machine learning method produced the best AUC, therefore machine learning was the best classifying method. This was also summarized in Table 3.6.
Table 3.6 SVM vs Multiple Regression Results

<table>
<thead>
<tr>
<th>Features Multi Regression</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR Contrast Energy</td>
<td>70</td>
<td>100</td>
<td>0.7</td>
<td>0.86</td>
</tr>
<tr>
<td>MR Min Contrast Energy</td>
<td>71.43</td>
<td>95</td>
<td>0.6643</td>
<td>0.887</td>
</tr>
<tr>
<td>MR Min Contrast Homogeneity</td>
<td>71.43</td>
<td>88.33</td>
<td>0.5976</td>
<td>0.858</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Machine Learning</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Youden Index</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min Contrast Energy (Gauss) n=13</td>
<td>100</td>
<td>85.7</td>
<td>0.857</td>
<td>0.97619</td>
</tr>
<tr>
<td>Min Contrast Homogeneity (Gauss) n=13</td>
<td>83.3</td>
<td>100</td>
<td>0.833</td>
<td>1.00</td>
</tr>
<tr>
<td>Contrast Energy n=130</td>
<td>96.7</td>
<td>84.3</td>
<td>0.961</td>
<td>0.81</td>
</tr>
<tr>
<td>Min Contrast Energy (Cubic) n=130</td>
<td>78.3</td>
<td>92.9</td>
<td>0.951</td>
<td>0.912</td>
</tr>
<tr>
<td>Min Contrast Homogeneity (Cubic) n=130</td>
<td>75</td>
<td>87.1</td>
<td>90.1</td>
<td>0.621</td>
</tr>
</tbody>
</table>

The cubic Support Vector Machine and the fine Gaussian support vector machine greatly increased classification of EBUS images. Sensitivity, Specificity, AUC and Youden Index were better with the machine learning method than the multiple regression method, as seen in Table 3.6 in the bolded regions.

Cross-validation used in both sample sizes caused overfitting for the support vector machine method. The purpose for the cross validation was to increase the number of samples that you train and test with, however the data that you train and test with was the subset of the same data we were investigating. Thus, the cross validation would inherently produce better results due to overfitting the classifier. The ideal method would
be to use the hold-out method to train and test the data. The hold-out method uses a percentage of the data set for training and then tests on the remainder of the dataset for accuracy of the model. Limited data sample did not allow testing the hold-out method, with more samples acquired in the future I would be able to test this method for training purposes.

3.6 Comparison of the N=13 to N=130 Methods

Although the EBUS images were confirmed clinically as malignant or benign, we don’t know the extent of the imaged lesions. Therefore, selecting multiple ROIs can lead to inaccuracies when it comes to classifying the lesion for that particular ROI. This poses a problem for the true accuracy of the study because that particular ROI may not be malignant or benign. Increasing the number would increase the error in characterizing the sampling (in addition to artificially increasing the experimental data size). When performing machine learning or predictive statistics it is critical to know the exact class of the training data for the training purposes. Thus, the ideal situation would be to create the ROI in the region that was biopsied by the clinician and proven by the histology analysis.

A t-test for a Hypothesized Mean is the method that was used to assess how well the 10 ROIs represents the whole lesion. This t-test was implemented by using the full lesion as the hypothesized mean and that same lesion split into 10 different samples are representative of the lesion (with the matching means). The purpose behind this test is also to find out if that specific ROI can be considered as malignant or benign, as represented mathematically below:

\[
H_0 = m_1 = m_2 \tag{3.1}
\]

\[
H_1 = m_1 \neq m_2 \tag{3.2}
\]
where $H_0$ represents the null hypothesis and $H_1$ represents the alternative hypothesis, $m_1$ and $m_2$ are the means of the samples being compared (ROI vs the whole lesion).

**Table 3.7 Malignant t-test**

<table>
<thead>
<tr>
<th>Row</th>
<th>h</th>
<th>p</th>
<th>ci_low</th>
<th>ci_high</th>
<th>tstat</th>
<th>df</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>3.69E-05</td>
<td>44.582</td>
<td>61.932</td>
<td>-7.501</td>
<td>9</td>
<td>12.126</td>
</tr>
<tr>
<td>STD</td>
<td>0</td>
<td>0.829</td>
<td>1.754</td>
<td>3.135</td>
<td>0.221</td>
<td>9</td>
<td>0.965</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>6.43E-05</td>
<td>42.739</td>
<td>61.460</td>
<td>-6.984</td>
<td>9</td>
<td>13.084</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>0.0015</td>
<td>37.189</td>
<td>55.810</td>
<td>-4.495</td>
<td>9</td>
<td>13.014</td>
</tr>
<tr>
<td>Max</td>
<td>1</td>
<td>0.000279</td>
<td>85.693</td>
<td>107.906</td>
<td>-5.743</td>
<td>9</td>
<td>15.526</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>0.00788</td>
<td>13.343</td>
<td>25.856</td>
<td>-3.398</td>
<td>9</td>
<td>8.745</td>
</tr>
<tr>
<td>Max_Min</td>
<td>1</td>
<td>6.88E-05</td>
<td>71.057</td>
<td>83.342</td>
<td>-6.923</td>
<td>9</td>
<td>8.586</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0</td>
<td>0.347</td>
<td>1.764</td>
<td>8.638</td>
<td>0.991</td>
<td>9</td>
<td>4.804</td>
</tr>
<tr>
<td>Skewness</td>
<td>1</td>
<td>0.0340</td>
<td>-0.207</td>
<td>0.451</td>
<td>2.496</td>
<td>9</td>
<td>0.460</td>
</tr>
<tr>
<td>Contrast</td>
<td>1</td>
<td>2.39E-05</td>
<td>678.941</td>
<td>739.159</td>
<td>7.924</td>
<td>9</td>
<td>42.089</td>
</tr>
<tr>
<td>Correlation</td>
<td>0</td>
<td>0.0538</td>
<td>-0.034</td>
<td>0.0109</td>
<td>2.216</td>
<td>9</td>
<td>0.0314</td>
</tr>
<tr>
<td>Energy</td>
<td>1</td>
<td>7.99E-07</td>
<td>0.000249</td>
<td>0.0002</td>
<td>-11.947</td>
<td>9</td>
<td>9.33E-06</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>1</td>
<td>0.00524</td>
<td>0.098</td>
<td>0.1055</td>
<td>-3.658</td>
<td>9</td>
<td>0.004</td>
</tr>
<tr>
<td>Entropy</td>
<td>1</td>
<td>5.67E-06</td>
<td>5.680</td>
<td>5.851</td>
<td>-9.460</td>
<td>9</td>
<td>0.119</td>
</tr>
<tr>
<td>Height</td>
<td>1</td>
<td>4.3E-06</td>
<td>109.821</td>
<td>133.178</td>
<td>9.782</td>
<td>9</td>
<td>16.324</td>
</tr>
<tr>
<td>Width</td>
<td>1</td>
<td>9.24E-07</td>
<td>91.533</td>
<td>102.666</td>
<td>-11.745</td>
<td>9</td>
<td>7.781</td>
</tr>
<tr>
<td>H_W</td>
<td>1</td>
<td>5.05E-06</td>
<td>1.098</td>
<td>1.428</td>
<td>9.592</td>
<td>9</td>
<td>0.230</td>
</tr>
</tbody>
</table>

Table 3.7 and Table 3.8 show statistics on how well the multiple ROI sampling represents the whole lesion. In this t-test for a hypothesized mean if a feature rejects the null hypothesis it means that it is not part of the population (lesion). From the malignant t-test (shown in Table 3.7) it can be observed from the bolded regions that Correlation, Standard Deviation and Kurtosis fall within the same population for both datasets. These
features are the only features that would be allowed for the classification in the malignant images for N=130 samples. Unfortunately, these three features were not suitable for classification. Therefore, by spreading out the ROI’s over the malignant lesion this increases the amount of error in the method and should not be used on the malignant images.

Table 3.8 Benign t-test

<table>
<thead>
<tr>
<th>Row</th>
<th>h</th>
<th>p</th>
<th>ci_low</th>
<th>ci_high</th>
<th>tstat</th>
<th>df</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>0.0135</td>
<td>59.94</td>
<td>74.265</td>
<td>3.061</td>
<td>9</td>
<td>10.013</td>
</tr>
<tr>
<td>STD</td>
<td>1</td>
<td>3.07E-5</td>
<td>2.547</td>
<td>4.025</td>
<td>-7.677</td>
<td>9</td>
<td>1.033</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>0.00249</td>
<td>58.957</td>
<td>73.843</td>
<td>4.148</td>
<td>9</td>
<td>10.405</td>
</tr>
<tr>
<td>Mode</td>
<td>1</td>
<td>0.0101</td>
<td>46.81</td>
<td>68.989</td>
<td>3.243</td>
<td>9</td>
<td>15.502</td>
</tr>
<tr>
<td>Max</td>
<td>0</td>
<td>0.609</td>
<td>103.194</td>
<td>128.805</td>
<td>0.53</td>
<td>9</td>
<td>17.901</td>
</tr>
<tr>
<td>Min</td>
<td>1</td>
<td>0.00919</td>
<td>27.065</td>
<td>44.735</td>
<td>3.303</td>
<td>9</td>
<td>12.351</td>
</tr>
<tr>
<td>Max_Min</td>
<td>0</td>
<td>0.0905</td>
<td>68.285</td>
<td>91.914</td>
<td>-1.896</td>
<td>9</td>
<td>16.516</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0</td>
<td>0.399</td>
<td>2.119</td>
<td>4.967</td>
<td>0.886</td>
<td>9</td>
<td>1.991</td>
</tr>
<tr>
<td>Skewness</td>
<td>0</td>
<td>0.472</td>
<td>-0.456</td>
<td>0.28</td>
<td>0.751</td>
<td>9</td>
<td>0.515</td>
</tr>
<tr>
<td>Correlation</td>
<td>1</td>
<td>8.81E-4</td>
<td>-0.0672</td>
<td>0.0351</td>
<td>4.872</td>
<td>9</td>
<td>0.0715</td>
</tr>
<tr>
<td>Energy</td>
<td>1</td>
<td>6.02E-12</td>
<td>0.000837</td>
<td>0.000879</td>
<td>45.47</td>
<td>9</td>
<td>0.0000294</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>1</td>
<td>1.67E-07</td>
<td>0.147</td>
<td>0.1621</td>
<td>14.34</td>
<td>9</td>
<td>0.0101</td>
</tr>
<tr>
<td>Entropy</td>
<td>0</td>
<td>0.0661</td>
<td>5.43</td>
<td>5.922</td>
<td>-2.091</td>
<td>9</td>
<td>0.341</td>
</tr>
<tr>
<td>Height</td>
<td>1</td>
<td>1.06E-12</td>
<td>36.065</td>
<td>41.334</td>
<td>-55.204</td>
<td>9</td>
<td>3.683</td>
</tr>
<tr>
<td>Width</td>
<td>0</td>
<td>0.283</td>
<td>88.196</td>
<td>109.203</td>
<td>-1.142</td>
<td>9</td>
<td>14.682</td>
</tr>
<tr>
<td>HW</td>
<td>1</td>
<td>8.36E-09</td>
<td>0.338</td>
<td>0.47</td>
<td>-20.188</td>
<td>9</td>
<td>0.0919</td>
</tr>
</tbody>
</table>

In Table 3.8 it is shown in bold that the features of Max, Max-Min, Kurtosis, Skewness, Entropy and Width do not reject the null hypothesis. In Table 3.7 it is shown
that only the features STD, Kurtosis and Correlation do not reject the null hypothesis. Neither table shows the features Min, Energy, Homogeneity or Contrast as accepting the null hypothesis, meaning the N=130 sampling is not a viable, as expected since one introduces additional data artificially. Thus, one needs to use the N=13 sampling approach.
4 CONCLUSION AND FUTURE DIRECTIONS

The diagnosis of EBUS images is a complex problem that requires the analysis of many variables to find the optimal parameters for classification. The classification of EBUS images can’t be simply done by comparing one parameter alone. Although the comparison of histogram features provided a good start at classification, it did not provide a strong discrimination power (sensitivity and specificity) between benign and malignant lesions. As a first step, the multiple regression was used to combine the parameters to achieve an improved classification results. Multiple regression also provided the optimal parameters to combine in the machine learning algorithm. The machine learning method used the features that were discovered with histogram and texture analysis and used a hyper plane to separate and classify the data. The support vector machine with the Gaussian kernel gave the best classification with the following features: Minimum pixel value, Contrast and Energy. A relatively high classification accuracy (0.98) was obtained with the machine learning method. However, the sample size of 13 was small, thus the results presented in this thesis cannot be representative of more general cases that would involve substantially higher number of samples.

To have a practical implementation of statistical analysis of EBUS images to characterize lesions, more EBUS images need to be analyzed in the future, so that there will be better accuracy and representation of the larger population. The ROI selection and classification algorithms can be implemented at the operating room, near real time so that these quantitative imaging parameters or features can be visualized at the monitor next to EBUS images. Thus, clinicians will have an opportunity to see the raw images as well as quantitative metrics during the biopsy sampling. If the histology and the image processing results correlate, then image-guided biopsy sampling could be established, which would speed up the clinical biopsy analysis as well. This step would allow
physicians to make more informed decisions for the biopsy in the lung and to have an initial idea of what is needed to treat the patient before they get the confirmed histology results. The physician could already have a quantitative imaging approach established as an alternative to waiting for the histology analysis so that earlier intervention could be realized. Quantitative image analysis could initially pinpoint the areas of interest for the physician to biopsy the most representative sites so that histopathology outcome could improve as well.
Bibliography


