IMAGE ANALYSIS TECHNIQUES FOR CLASSIFICATION OF PULMONARY DISEASE IN CATTLE

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Biomedical Engineering

University of Saskatchewan

Saskatoon

By

C. Denise Miller

Keywords: image processing, medical image analysis, histopathology

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biomedical Engineering

University of Saskatchewan

Saskatoon, Saskatchewan

S7N 5A9

ABSTRACT

Histologic analysis of tissue samples is often a critical step in the diagnosis of disease. However, this type of assessment is inherently subjective, and consequently a high degree of variability may occur between results produced by different pathologists. Histologic analysis is also a very time-consuming task for pathologists. Computer-based quantitative analysis of tissue samples shows promise for both reducing the subjectivity of traditional manual tissue assessments, as well as potentially reducing the time required to analyze each sample.

The objective of this thesis project was to investigate image processing techniques and to develop software which could be used as a diagnostic aid in pathology assessments of cattle lung tissue samples. The software examines digital images of tissue samples, identifying and highlighting the presence of a set of features that indicate disease, and that can be used to distinguish various pulmonary diseases from one another. The output of the software is a series of segmented images with relevant disease indicators highlighted, and measurements quantifying the occurrence of these features within the tissue samples. Results of the software analysis of a set of 50 cattle lung tissue samples were compared to the detailed manual analysis of these samples by a pathology expert.

The combination of image analysis techniques implemented in the thesis software shows potential. Detection of each of the disease indicators is successful to some extent, and in some cases the analysis results are extremely good. There is a large difference in accuracy rates for identification of the set of disease indicators, however, with sensitivity values ranging from a high of 94.8% to a low of 22.6%. This wide variation in result scores is partially due to limitations of the methodology used to determine accuracy.

ACKNOWLEDGMENTS

I would like to extend thanks to my supervisor, Dr. Tony Kusalik, for his guidance and support during this work. I am extremely grateful to Dr. Gary Wobeser for his unending patience in instructing me in histologic analysis, and for the time and effort he devoted to the manual analysis of the test images for comparison to the software results. My thanks are also extended to Dr. Mark Eramian for his ideas and guidance on image processing techniques. A special thank-you is extended to Dr. Greg Appleyard for suggesting this work, for his help with the project, and for being an enthusiastic champion of the project throughout.

This work was partially funded by a CFI Innovation Fund grant, and a Province of Saskatchewan Innovation and Science Fund grant. Financial assistance for the project was provided through a University Graduate Scholarship from the University of Saskatchewan.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The Histopathology Challenge. Accurate diagnosis of disease, and its underlying cause, typically requires a combination of careful clinical observation, a thorough case history, and diagnostic testing of biological samples. Visual assessment of images, including histologic analysis of tissue samples, is one key area of diagnostic testing. Unlike other laboratory tests which are fully quantifiable, histologic analysis is inherently subjective, and consequently prone to a high degree of variance in results produced by different pathologists. Numerous studies of inter-observer variation in histopathology assessments have been published in recent pathology literature, and in some cases have shown that not only is there a notable lack of consensus amongst pathologists examining the same samples, but also that the same pathologist can produce different results when examining the same specimen on different occasions [1]. When rating the level of agreement between pathologists in recognizing the presence, absence, or degree of histological criteria, various studies have found only fair to moderate concordance [2].

Histologic analysis is also a time-consuming task for pathologists. Using the level of magnification required in order to view details in a tissue sample, only a small portion of the sample is visible in a microscope's field-of-view at one time. The amount of time required for careful evaluation of disease indicators in the tissue within each field-of-view is significant, and, at the required magnification level, complete analysis of the full tissue sample would require examination of hundreds of such field-of-view images. Since this would be time-prohibitive, only a few areas of each sample are selected for close inspection.

Computer-based quantitative analysis of tissue samples shows promise for reducing the subjectivity of histologic assessment, potentially yielding more consistent results. Automated analysis software could also reduce the amount of pathologists' time required for these analyses. By creating digital images from standard tissue sample slides, image processing and computer vision techniques can be used to automate parts of manual histopathology analyses. Analysis software that provides a full diagnosis of the disease state present in each tissue sample would,

of course, deliver the most benefit. However, significant time savings could also be realized through the use of software which aids the pathologist in examining samples, even if the final diagnosis is left to the pathology expert. For example, computer-based analysis software could examine full tissue samples, highlighting to the pathologist areas of the samples which are most indicative of disease. This analysis could potentially reduce oversight errors resulting from the manual selection of particular regions of the tissue sample for close inspection. Secondly, automated analysis could quickly separate routine cases from those that are more complicated. Diagnosis of the high percentage of "typical" cases of easily identifiable diseases could be streamlined, and pathology expertise could be focused against the remaining minority of more challenging samples identified within the screening, those that truly require more detailed examination.

Computer Vision and Image Processing. The goal of computer vision is equivalent to that of human vision: the perception and understanding of images. Spectral, textural, and contextual features are fundamental elements used in human interpretation of images [3], and these same elements play an essential role in computerized interpretation of images. However, the manner in which this information is processed is fundamentally different. In human vision, recognition of objects is a qualitative process. In machine vision, the process is predominantly quantitative.

Computer vision is generally divided into two levels of processing. Low-level image processing techniques operate on fundamental image data, such as arrays of image pixel intensities. At this level, little knowledge of the actual content of the image is used. In contrast, high-level image processing is based on knowledge of the image content. Pattern recognition and artificial intelligence techniques are often used as part of this level of processing. Often low- and high-level processing techniques are used iteratively, with low-level processing steps extracting information required for high-level processing, and high-level processing identifying gaps in image understanding where additional low-level processing may be beneficial. Low-level processing includes such steps as noise reduction, image compression, edge extraction, partial image segmentation, and object classification. High-level processing aims to interpret and understand the image contents, by applying such techniques as statistical pattern recognition. Some authors prefer to also distinguish an intermediate level of processing, which includes techniques used to obtain abstract information about image features, such as the existence of

certain shapes within the image, without attempting to apply any real world meaning to these objects [4].

Application of Image Processing Techniques to Biological Images. Automated image analysis has been successfully applied in numerous fields. However, computerized interpretation of biomedical images is a complicated problem. Statistical, structural and temporal variations of biological structures make automated analysis of these images difficult. Furthermore, existing feature extraction techniques may make simplifying assumptions which are unrealistic for biological images; for example, assuming that objects are rigid [5]. Even with these challenges, application of image processing techniques to biological images is an active area of research (as discussed further in section 3.4).

Goals of the Thesis Project. Although the ultimate goal of computerized image analysis of tissue samples would be a fully automated diagnosis based on the image contents, significant benefits could also be realized from software which determines the presence of disease-indicating features within the samples, highlights these areas for pathologists, and produces measurements of these relevant features. This quantitative data could then be used by pathologists to assist in arriving at a diagnosis.

The goal of this thesis work was to investigate image processing techniques and to develop image analysis software which could be used as such a diagnostic aid in pathology assessments of cattle lung tissue samples. The software created examines digital images of tissue samples, identifying and highlighting the presence of a certain set of features that indicate disease, and that can be used to distinguish various pulmonary diseases from one another. The output of the current software is a series of segmented images with relevant disease-indicating features highlighted, and measurements quantifying the occurrence of these features within the tissue samples. The software does not diagnose the disease states present, although this is possibly feasible as an extension to this work.

Structure of this Document. This thesis contains four main sections: Objectives, Background, Data & Methodology, Results, and Conclusions, Discussion & Future Work. The Objectives section clarifies the approach used for the thesis project, including the limitations of this work. In the Background section, basic information is presented that is critical to understanding the work presented. Within the Background section, subsection 3.1 provides an overview of the various pulmonary diseases represented in the tissue samples analyzed. This

section also provides definitions and examples of the subset of disease-indicating features that the project software detects and measures. Section 3.2 introduces some image processing concepts that are fundamental to the techniques applied in the image analysis software for this project. Subsection 3.3 provides a very brief overview of classification algorithms employed as part of the methodology used for this project. Finally, subsection 3.4 examines some applications of image processing techniques to the analysis of biological images, as cited in recent literature.

Within the Data & Methodology section, subsection 4.1 describes the tissue sample slides analyzed as part of this project, and also outlines the approach used to select regions of each sample for detailed analysis and comparison to manually generated results. Subsection 4.2 provides an overview of the equipment used to create digital images from the tissue sample data. Subsection 4.3 is the most substantial portion of this document, providing full details of the three different approaches used to locate disease-indicating features within the tissue sample images.

The Results section includes both the tissue sample analysis output from the project software, as well as the manual analysis results for comparison purposes, for some of the fifty tissue images included in this project. Measurements which compare the accuracy of the automated assessment to the manual analysis are provided within this section for all fifty cases. (Note: Although not included in this document, detailed automated and manual results for all remaining images are also available. See the Results section for details.)

Finally, the Conclusions, Discussion and Future Work section explores ways in which the current work could be improved and expanded, as well as explaining the results obtained.

CHAPTER 2
OBJECTIVES

The objective of this thesis project was to investigate image processing techniques, and to develop image analysis software which can locate and measure a specified subset of disease-indicating features within digital images of cattle lung tissue samples. Essentially, this work involved development of a prototype system capable of identifying features of interest within the tissue samples.

The approach for this work included:

- Determining a set of image processing techniques that can be used to successfully extract features of interest from digital images of cattle lung tissue samples; adapting and expanding these techniques as necessary.

- Developing software to implement these techniques. The software isolates features within the images that are indicative of disease, and which could distinguish different pulmonary diseases from one another. The software is capable of highlighting these features, and displaying highlighted versions of the images to a pathologist.

- The software produces statistics to summarize the presence or absence of the features of interest within each image.

The thesis work was subject to the following limitations:

- The thesis does not include an in-depth examination of any one image processing technique. Instead, the focus of the thesis work was to obtain an adequate level of understanding of numerous image processing techniques in order to determine their suitability for the analysis of the lung tissue sample images.

- The thesis does not include a comparison of image processing techniques. As part of the thesis work, various techniques were evaluated to determine their usefulness for analyzing the tissue samples. The thesis includes an explanation of all techniques which were used, and how each of these techniques contributed to the analysis of the tissue samples. The thesis does not include a comparative evaluation of techniques which were examined and rejected.

CHAPTER 3
BACKGROUND

3.1 Pulmonary Disease in Cattle

Bovine Respiratory Disease (BRD) is a significant cause of economic loss to the cattle industry, and has been identified as the primary health problem of feedlot cattle [6]. Annual loss of cattle due to respiratory disease is reported to be higher than deaths due to any other cause [6]. Consequently, early and accurate diagnosis and treatment of BRD outbreaks is of critical importance. With respiratory diseases, clinical observation is often the basis for initially determining treatment and preventative protocols [7]. However, when fatalities occur, histopathologic analysis often plays a key role in confirming the initial diagnosis, identifying the underlying cause of disease, and adjusting treatment and preventive measures for other cattle in the herd.

Within the broader category of BRD, specific pulmonary diseases include bronchitis, bronchiolitis, and numerous types of pneumonia (e.g. interstitial, focal, bronchopneumonia). Gross observation of the pattern of involvement of the lung is one essential step in distinguishing between the various pulmonary diseases. Histologic examination of lesions also plays an important role, allowing the pathologist to determine which disease indicators are present. Symptoms observable at the microscopic level include emphysema, atelectasis, edema, hemorrhage, infiltration of neutrophils and macrophages, necrosis, and fibrin.

The thesis software detects and measures the following subset of disease indicators: edema, infiltration of neutrophils and/or macrophages, presence of fibrin, and areas of necrosis. Also, although the software does not directly identify atelectasis, measurements produced by the software could be used to decide whether this condition is present.

3.1.1 Edema

Pulmonary edema is fluid accumulation within the alveoli of the lungs. When tissue samples are stained using standard hematoxylin and eosin (H&E) histology stain, fluid will

appear a light to darker pink colour, depending on the amount of protein content (Figure 3.1). Edema fluid will contain air bubbles of various sizes.

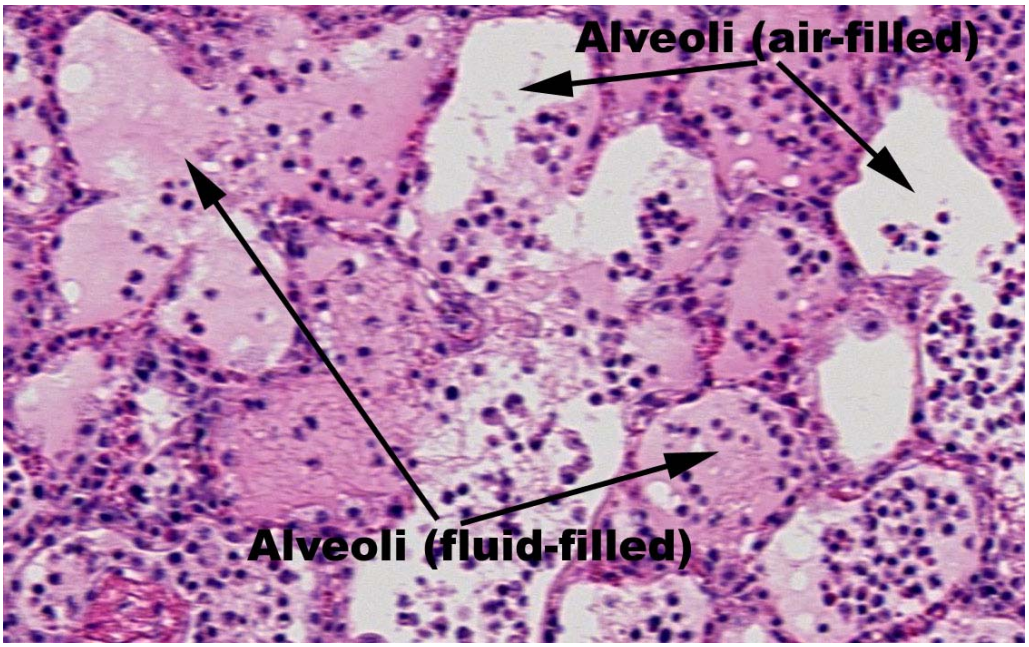


Figure 3.1: Air-filled and fluid-filled alveoli (edema)

3.1.2 Areas of inflammatory cells (infiltration of neutrophils and macrophages)

Sites of inflammation are characterized by an infiltration of leukocytes, including neutrophils and macrophages. Sixty to seventy percent of all blood leukocytes are polymorphonuclear granulocytes, of which neutrophils are the most common variety [8]. Neutrophils use phagocytosis to ingest invading organisms, such as bacteria. The presence of neutrophils is a key indicator to pathologists that bacteria may be present, although in the very early stages of bacterial pneumonia macrophages are likely to be the responding cell instead [9]. Vast numbers of neutrophils will be visible in lung tissue samples for cases of suppurative pneumonia, for example.

Twenty percent of blood leukocytes are lymphocytes, including macrophages [8]. These non-specific leukocytes also use phagocytosis to eliminate invading particles. In the lung, all macrophages are bone marrow derived (B-cells). When foreign substances invade the lung, the first response is by macrophages, followed by a neutrophil response [10]. Although some number of macrophages will always be present, a high concentration of these cells in a small

area is abnormal. Large numbers of macrophages will be present in alveolar walls in cases of interstitial pneumonia, for example.

When tissue samples are prepared using H&E stain, the nuclei of neutrophils and macrophages will appear a dark purple / blue colour, and cytoplasm will appear pink (Figure 3.2). At the level of magnification used for the thesis work, it is mainly the nuclei which are the noticeable feature of these cells. Nuclei of neutrophils are multi-lobed; those of macrophages are less dense than nuclei of neutrophils, with a "grainier" appearance.

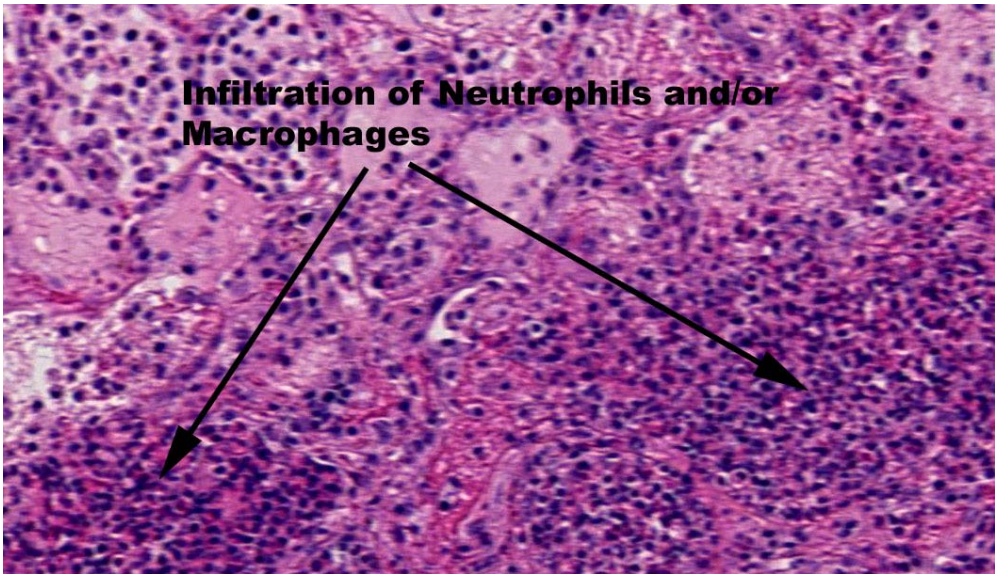


Figure 3.2: Areas of inflammatory cells (infiltration of neutrophils and/or macrophages)

3.1.3 Fibrin

Fibrin is a polymerized protein substance which is involved in the formation of plugs or clots within injured tissue. With H&E staining, fibrin will appear light to dark pink. Depending on the amount of aggregation which has occurred, a mesh-like structure may be visible microscopically. In later stages, fibrin appears as a dense, dark pink area (Figure 3.3). Fibrin is a key feature in fibrinous bronchopneumonia, for example.

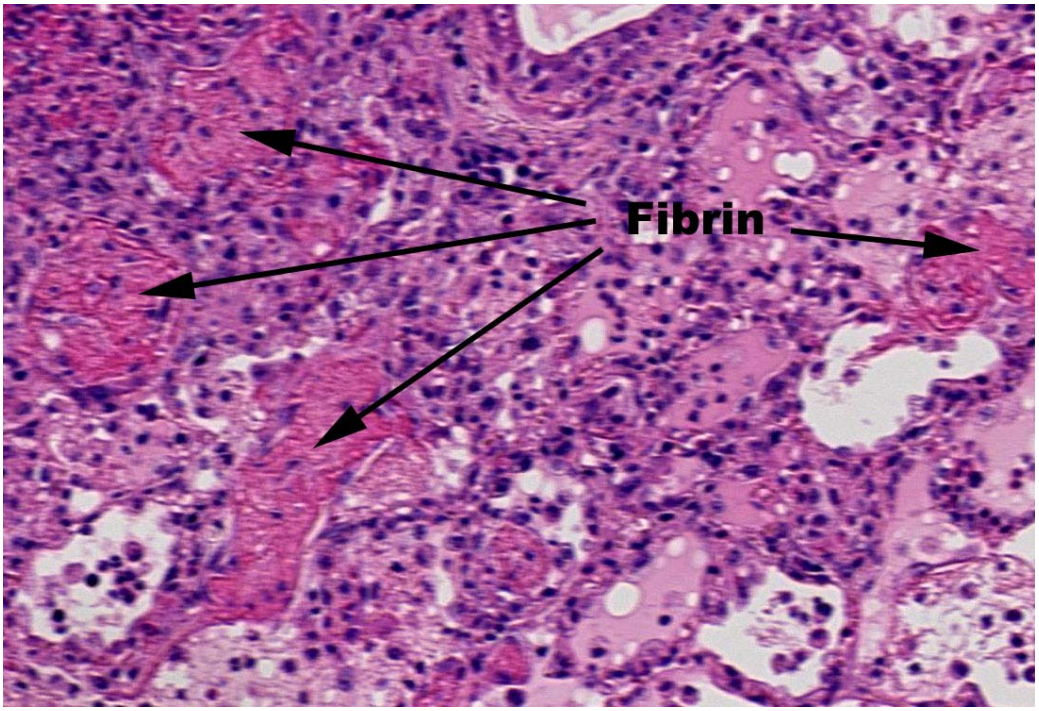


Figure 3.3: Areas of fibrin

3.1.4 Necrosis

Necrosis is tissue death. Using H&E stain, necrotic tissue will appear a solid pink colour, with no visible nuclei (Figure 3.4). Neutrophils are attracted to sites of necrosis, and areas of neutrophils may be visible surrounding necrotic areas. Necrosis may be present as a feature of a number of different pulmonary diseases, including suppurative and fibrinous bronchopneumonia, and broncho-interstitial pneumonia.

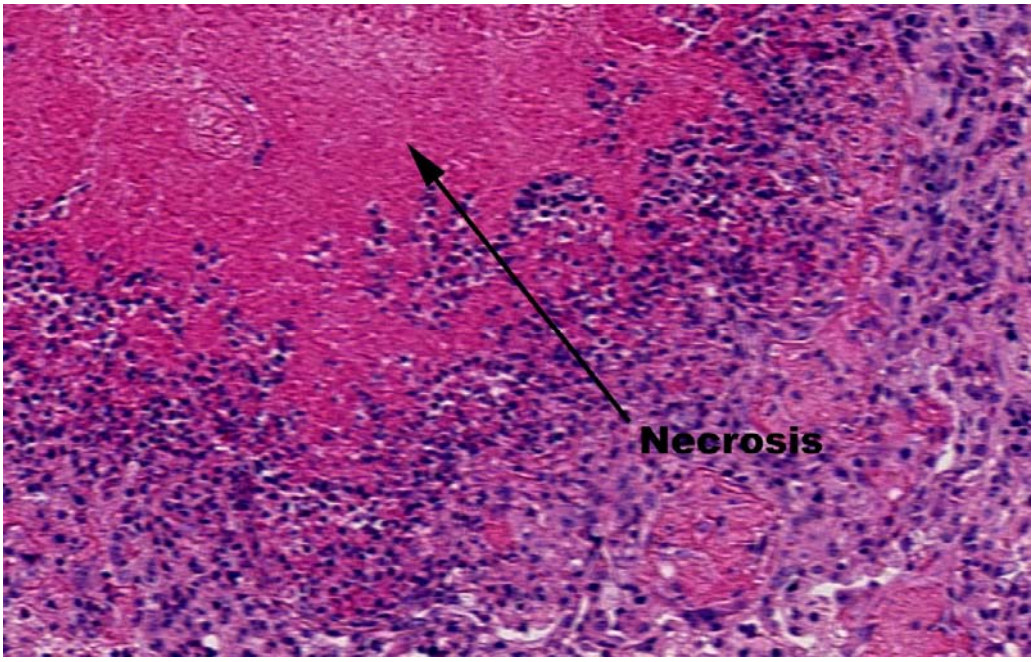


Figure 3.4: Area of necrosis

3.1.5 Atelectasis

Atelectasis is a condition in which alveoli have collapsed, resulting in decreased pulmonary air capacity. Microscopically, alveoli will appear to have closely placed walls, with small or non-existent lumen (Figure 3.5).

Although the thesis software does not directly detect atelectasis, overall alveolar space in each tissue sample is measured. An abnormally low value for alveolar area in any tissue sample could indicate atelectasis.

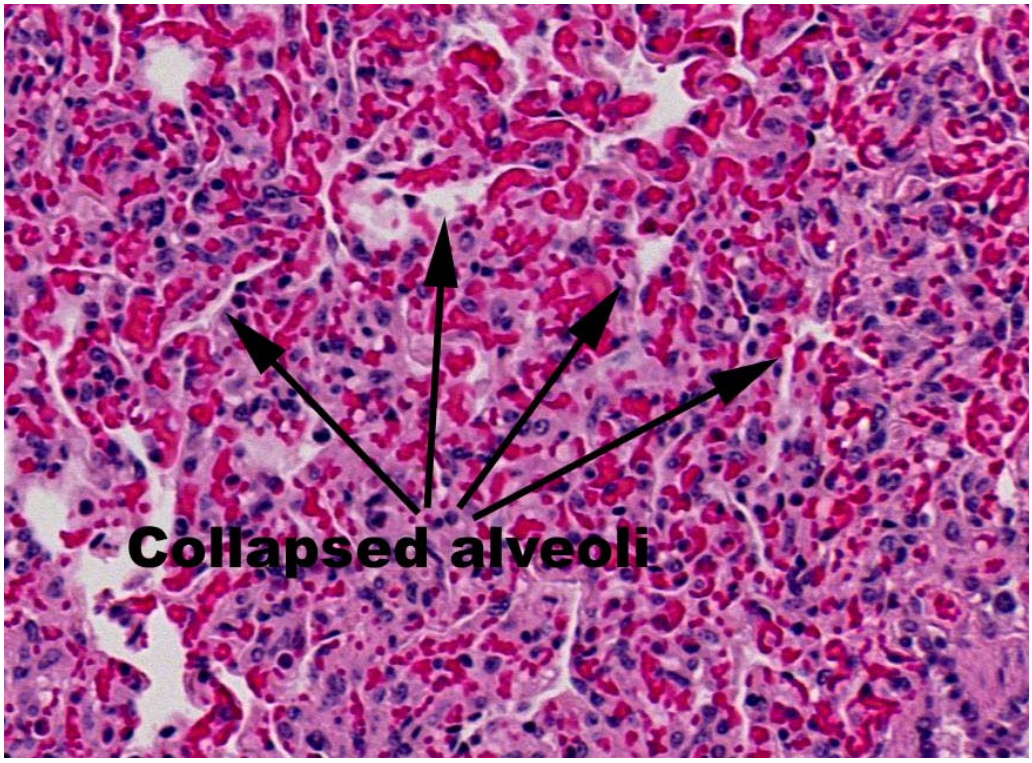


Figure 3.5: Atelectasis

3.2 Image Processing Background

3.2.1 Image Representation

An image captured by a sensor is a continuous function, f(x,y), of two co-ordinates in the plane. Digitization is the process through which this function is sampled into a matrix of M rows by N columns of individual picture cells, or pixels. Image quantization assigns to each sample a discrete integer value (or values), thus translating the continuous range of the image function into k intervals, where k is the quantization number. The internal format of a digital image is thus a three-dimensional array in which the x and y co-ordinates describe the pixel position, and the z co-ordinate contains the pixel's intensity value. In greyscale images, the z co-ordinate contains a single value; in colour images, the z co-ordinate contains a vector of intensity values, with one value for each colour component.

R:206 G:189 B:202	R:201 G:175 B:197	R:190 G:158 B:188	R:187 G:149 B:191	R:182 G:137 B:189	R:172 G:122 B:173	R:172 G:120 B:179	R:172 G:110 B:168
R:206 G:181 B:200	R:199 G:172 B:190	R:196 G:157 B:182	R:189 G:149 B:177	R:177 G:133 B:188	R:166 G:129 B:183	R:163 G:120 B:172	R:165 G:110 B:166
R:204 G:186 B:219	R:196 G:166 B:193	R:189 G:157 B:199	R:183 G:155 B:186	R:176 G:131 B:185	R:175 G:123 B:175	R:163 G:118 B:172	R:159 G:105 B:176
R:213 G:194 B:218	R:194 G:180 B:206	R:190 G:163 B:199	R:189 G:148 B:192	R:185 G:134 B:196	R:182 G:134 B:191	R:168 G:118 B:179	R:162 G:109 B:167
R:217 G:211 B:207	R:208 G:193 B:214	R:207 G:184 B:205	R:194 G:164 B:207	R:190 G:154 B:195	R:184 G:142 B:188	R:181 G:132 B:175	R:174 G:122 B:173
R:226 G:224 B:216	R:216 G:218 B:201	R:211 G:201 B:203	R:205 G:182 B:199	R:199 G:163 B:195	R:197 G:154 B:188	R:181 G:141 B:193	R:174 G:123 B:182
R:231 G:219 B:219	R:229 G:218 B:219	R:231 G:216 B:212	R:219 G:204 B:212	R:214 G:185 B:194	R:206 G:174 B:214	R:191 G:153 B:194	R:179 G:145 B:196
R:229 G:223 B:218	R:231 G:224 B:216	R:234 G:221 B:219	R:229 G:213 B:214	R:224 G:210 B:212	R:206 G:185 B:203	R:199 G:175 B:205	R:190 G:165 B:186

Figure 3.6: Part of a digital image, enlarged to show individual pixels, with red-green-blue intensity values superimposed on each pixel

3.2.2 Connectivity

When examining pixel relationships in image processing, one key concept is that of connectivity. On a two-dimensional rectangular grid there are two possibilities for defining neighbouring pixels. Pixels can be considered neighbours if they share a joint edge (4-connectivity. See Figure 3.7), or if they have at least one common corner (8-connectivity. See Figure 3.8).

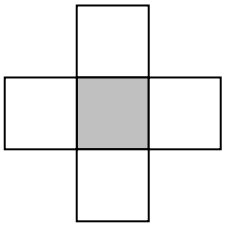


Figure 3.7: 4-connected neighbours of the centre pixel

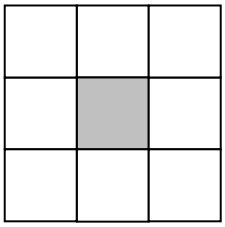


Figure 3.8: 8-connected neighbours of the centre pixel

3.2.3 RGB and HSV Colour Models

Colour conveys a significant amount of information, and is an important component in human visual perception of images. However, historically colour has not been used in digital image processing. Traditional use of greyscale images in image processing has been primarily motivated by the greater storage costs for multi-spectral image data, and the greater complexity in processing these images. In greyscale images, each pixel within an image is represented by a single number: the intensity value. In colour images, each pixel is represented by multiple numbers, each of which conveys the intensity in the independent colour components. The vast majority of colour models consist of three colour components, although colour models with additional spectral components are sometimes used [11].

The RGB colour model is the most common of the various colour models. In the RGB colour model, each pixel is represented by a three-dimensional vector, (r,g,b), which gives the pixel's intensity in each of the red, green and blue colour components. This representation of all colours as a combination of red, green and blue is consistent with the way colours are perceived by humans. The majority of devices used to acquire and display images utilize this colour model. The RGB colour model can be thought of as a three dimensional co-ordinatization of colour space (Figure 3.9). In this model, intensity numbers for each colour component range from 0 to the maximum quantization number. Eight bits per colour channel are most frequently used, yielding intensity values for each dimension which range from 0 to 255. In the RGB colour model, (0,0,0) represents black, (255,255,255) represents white, (255,0,0) represents pure red, etc. Every possible colour which can be represented using 24-bit colour (8 bits per channel) corresponds to some point within the colour cube.

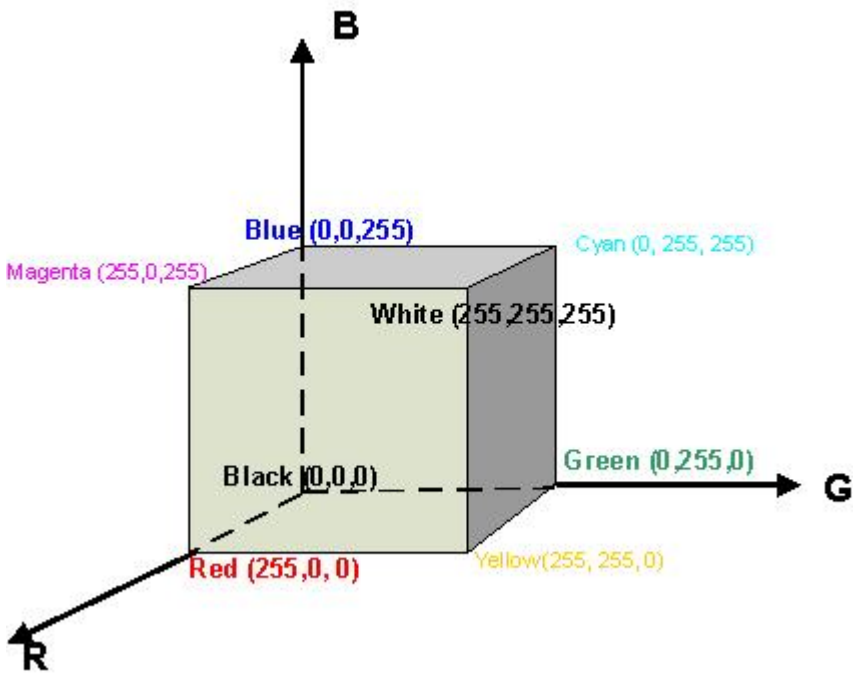


Figure 3.9: RGB colour cube

Another frequently used colour model is HSV (also known as HSI). In this colour model, intensity information is separated from the spectral information. Hue (H) refers to the perceived colour, such as purple, red etc. Saturation (S) measures the colour's dilution with white, which produces shades of the colour, such as light purple, dark purple etc. The third component is the value (V), or intensity, which describes the pixel's brightness. The HSV colour model forms a hexagonal cone shape, where hue is measured by the angle around the vertical axis to the colour point, saturation is measured as distance from the centre of the hexagonal plane to the point, and value is measured as the distance from the hexcone apex to the hexagonal plane containing the point, as shown in Figure 3.10.

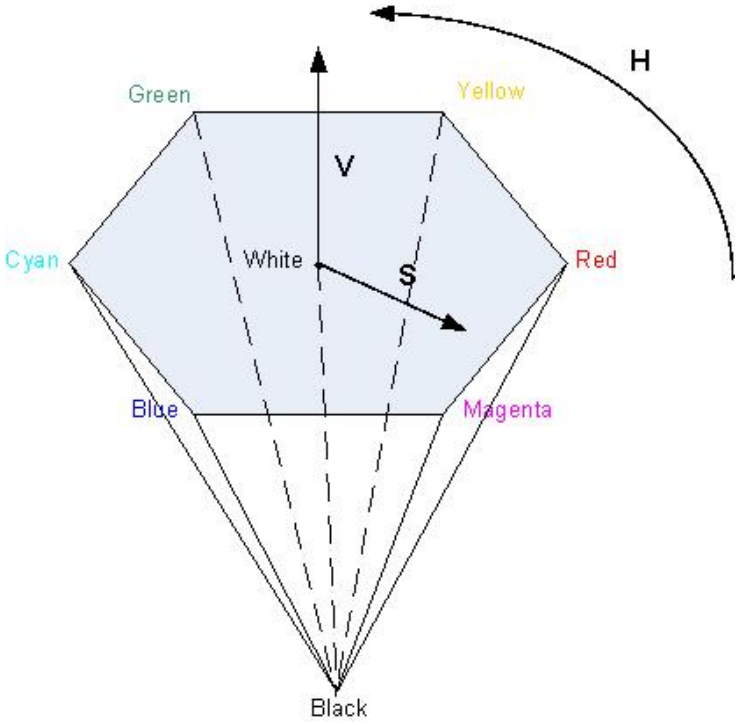


Figure 3.10: HSV hexcone

3.2.4 Image Segmentation

Image segmentation is one of the most important steps leading to analysis of image data. The goal of image segmentation is to divide an image into regions that correspond to objects in the real world. Meaningful objects are separated both from the image background and from one another. Total segmentation requires knowledge of the object domain, and is generally possible only for very simple tasks. In most cases, only partial segmentation is feasible. In partial segmentation, regions are meaningful but do not necessarily directly correspond to real world objects. One key advantage of partial segmentation is that it can be achieved without knowledge of the problem domain, since partial segmentation typically utilizes only low-level image data. Image segmentation techniques are classified as edge-based methods, which attempt to locate region boundaries, and region-based methods, which attempt to locate homogenous image areas.

3.2.5 Edges and Edge Detection

Edges are an important image feature, which are often used in image analysis to determine region boundaries. The significance of object boundaries cannot be overstated. According to Sonka et al., object boundaries are the most critical cue in the interpretation of intensity images [11]. Detection of edges is one of two main approaches in locating objects within images, the other being region-based methods.

Edges are a feature of individual pixels. An edge is defined as a pixel where the image intensity function changes abruptly. This change in intensity function can be described by a gradient that points in the direction of the maximum change in the function. Edges are expressed as vector variables with two components: magnitude and direction. The magnitude of an edge is the magnitude of the gradient at that point. The direction of an edge is defined as the direction of the gradient, rotated by -90 degrees.

Changes in continuous functions are described by derivatives. Since digital images are discrete, derivatives are approximated by differences. The first derivative of the image function should have an extremum at the position corresponding to an edge in the image; the second derivative should be zero at this same position. The majority of edge detection techniques are based either on approximations of the first derivative of the image function, or on the zero-crossings of the second derivative. A gradient function which approximates a first derivative is used in this thesis project to determine the location of edges.

Edge detection in multispectral images poses a unique set of challenges. In these images, an edge may appear in one colour component only, or in all colour components. Simple addition of gradients in all components is not useful, since gradients in two channels may end up cancelling each other [12]. More sophisticated techniques are required for edge detection in colour images. Although all images processed as part of this project are colour images, as a simplifying step images are converted to greyscale first, and edges are detected in these images instead.

3.2.6 Texture

3.2.6.1 Definition of texture

Texture is one of the fundamental elements used in the human interpretation of images, along with colour and context [3]. Although texture is an important and widely used image feature, there is no single, precise definition of texture in the image processing and computer vision literature. However, several properties of texture are generally accepted. Texture is a property of image areas, not of individual points. Texture involves the spatial distribution of tones, either grey-level or colour. When an area of an image has little variation in discrete tones, the dominant feature of that area is tone; when there is a wide variation in the discrete tones within an image area, the dominant feature is texture [3].

Textures consist of texture primitives, or texture elements (texels), the basic components which are repeated within the texture. A set of contiguous pixels which make up a blade of grass would be a texel within a close-up photograph of a lawn, for example. Often texels will represent physical world objects, although this is not required. An image region is perceived to have texture when the number of primitives within the area is large. If an image region contains only a few primitives, it is a group of countable objects that is perceived, instead of a texture [13]. The spatial relationship between texture primitives may be regular or random. This spatial relationship is referred to as the structure of the texture.

Texture analysis is primarily used for classification and texture segmentation. In texture classification, the goal is to determine the texture category to which an image region belongs. In texture segmentation, the aim is to determine boundaries between image regions of different textures. Texture classification and segmentation are often divided into the subprocesses of texture feature extraction, feature selection, and finally application of a classification or segmentation algorithm based on the selected features.

3.2.6.2 Describing texture

There are numerous means of describing texture, which Tuceryan and Jain categorize as statistical methods, geometrical methods, and model-based methods [13]. In this categorization, both geometrical and model-based texture description techniques use texture primitives. These types of texture description are suitable where texture primitives can be described using multiple

properties, such as tone and shape. In cases where texture primitive sizes are very small (comparable to pixel sizes, for example), statistical methods of texture description are more appropriate.

Statistical methods of texture description typically use either first-order or second-order spatial statistics. First-order statistics describe properties of individual pixel colours, without considering any interactions with neighbouring pixel values. These statistics measure the likelihood of observing a given grey-level or colour value at a given location in an image. Second-order statistics describe pixel colour relationships, and are properties of pairs of pixels. Second-order statistics measure the likelihood of observing two specific grey-level or colour values at a given distance and orientation from one another. Higher order statistics also exist, and are thought to be used as part of the human visual system [14]. Statistical methods of texture representation include autocorrelation features and co-occurrence matrices. Implementation of statistical methods of texture description is generally straightforward, but the key disadvantage is the vast amount of data involved, particularly for higher order statistics.

Geometrical methods of texture representation define textures as being composed of texture primitives. In these methods of texture description, either the texture elements are extracted and statistical properties of the texels measured, or placement rules that describe the arrangement of texels are used. Examples of geometrical methods of texture description include Vornoi tessellation features, and structural methods. Structural methods, which assume that texels are well-defined and arranged according to determinable placement rules, are of limited use however, since most image textures are not this regular.

Model-based methods of texture description involve construction of an image model that can be used to describe the texture. Examples of model-based methods include use of Markov random fields, fractals, and Fourier domain filtering. In the fractal method, for example, fractal functions are used as the texture models, with fractal dimension used as the feature for texture segmentation.

3.2.7 Co-occurrence Matrices

Co-occurrence matrices, also known as greylevel spatial dependence matrices, are one of the most widely used texture analysis tools. This form of statistical texture description specifies the grey-level spatial dependencies within a texture, computed at various angles and distances.

Co-occurrence matrices are advantageous in that they work well as texture descriptions for a large variety of textures, and provide good results when distinguishing between textures. Disadvantages of co-occurrence matrices include the fact that they are computationally expensive to construct, and have high memory requirements.

Each co-occurrence matrix summarizes the frequencies with which various grey-levels occur at a specific distance and specific angle from one another within an image region. Matrix $P_{0,1}$ summarizes the frequencies with which grey-levels occur at angle 0 degrees and distance 1 from one another, for example. Each matrix element contains the frequency for two specific grey-level values. For example, element $P_{0,1}$ (x,y) provides the frequency with which greylevels x and y appear at a distance of 1 in direction 0 degrees from one another. The dimensions of co-occurrence matrices will be the number of grey-levels within the image. In order to limit the size of co-occurrence matrices, it is necessary to restrict the number of grey-levels within the images. Co-occurrence matrices for a given distance are constructed for four angular directions: 0, 45, 90, 135 degrees. Using the most common definition of co-occurrence matrices, no distinction is made between the direction from grey-level x to grey-level y vs. the direction from grey-level y to grey-level x. As a result, it is not necessary to construct the matrices for the remaining directions [11].

As an example, consider the following image region which contains four different grey-level values:

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

For this image region, pixels with grey-levels 0 and 0 appear at a distance 1 and angle 0 degrees from one another a total of 4 times, so $P_{0,1}$ (0,0) = 4. Grey-levels 0 and 1 appear at distance 1 and angle 0 degrees a total of 2 times, so $P_{0,1}$ (0,1) = 2. By symmetry of the co-occurrence matrix, $P_{0,1}$ (1,0) is also 2. Grey-levels 0 and 2 appear at distance 1 and angle 0 a total of 1 time, yielding $P_{0,1}$ (0,2) = $P_{0,1}$ (2,0) = 1, and grey-levels 0 and 3 do not appear at distance 1 from one another, so $P_{0,1}$ (0,3) = $P_{0,1}$ (3,0) = 0. The full co-occurrence matrix $P_{0,1}$ is shown below.

$$\text{Co-occurrence matrix } P_{0,1} = \begin{vmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{vmatrix}$$

Co-occurrence matrices for this image region for distance 1 and the remaining three directions are shown below.

$$P_{45,1} = \begin{vmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix} \quad P_{90,1} = \begin{vmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{vmatrix} \quad P_{135,1} = \begin{vmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{vmatrix}$$

(Example reproduced from [11].)

Two key decisions in constructing co-occurrence matrices are the size of the image region to be described by the matrices, and the distance to use. If selected image regions are too small, the resulting matrices will not contain enough textural information. If image regions are too large, the matrices may summarize data from objects belonging to several categories [3]. As for selecting the distance, there is no well-established method for selecting the appropriate displacement to use for an image [13].

The spatial statistics provided by co-occurrence matrices are not utilized directly, but rather are used to compute various textural features. These features contain information about the textural characteristics of the image. Some of the co-occurrence matrix features, including contrast and correlation, have reasonable correspondence to characteristics detected in human perception. However, there is not a direct correspondence of many of these features to specific textural characteristics which are visible to the human eye. In the original paper describing the use of greylevel co-occurrence matrices as a texture analysis tool, Haralick et al. listed a total of fourteen texture features that could be calculated from these matrices [3]. Of these, the most commonly used co-occurrence matrix features are: energy (also referred to as angular second momentum), entropy, contrast, homogeneity and correlation [13]. For this project, the co-occurrence matrix features contrast, correlation, homogeneity and energy are used. These features are defined as follows:

- Contrast measures the amount of local variation in an image. Images with large edge magnitudes will have high contrast values.

$$\text{Contrast} = \sum_{a,b} |a-b|^{\kappa} P^{\lambda}_{\varphi,d}(a,b) \tag{3.1}$$

where typically $\kappa = 2$, $\lambda = 1$.

- Correlation measures linear dependence in an image. Linear structures in a given direction will result in large correlation values in this direction.

$$\text{Correlation} = \frac{\sum_{a,b}\left[(ab)P_{\varphi,d}(a,b)\right] - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{3.2}$$

where μ_x, μ_y are means and σ_x, σ_y are standard deviations:

$$\mu_x = \sum_a a \sum_b P_{\varphi,d}(a,b) \tag{3.3}$$

$$\mu_y = \sum_b b \sum_a P_{\varphi,d}(a,b) \tag{3.4}$$

$$\sigma_x = \sum_a (a-\mu_x)^2 \sum_b P_{\varphi,d}(a,b) \tag{3.5}$$

$$\sigma_y = \sum_b (b-\mu_y)^2 \sum_a P_{\varphi,d}(a,b) \tag{3.6}$$

- Homogeneity, as its name suggests, measures the uniformity of the image.

$$\text{Homogeneity} = \sum_{a,b} \frac{P_{\varphi,d}(a,b)}{1+|a-b|} \tag{3.7}$$

- Energy, or angular second momentum, also measures homogeneity of an image. If the image is homogenous, there are few grey-level transitions, so the energy measurement for the image will be large.

$$\text{Energy} = \sum_{a,b} P^2_{\varphi,d}(a,b) \tag{3.8}$$

Features calculated from any one co-occurrence matrix will be angularly dependent. To create a rotationally invariant texture description based on co-occurrence matrices, the matrices should be constructed for the four angular directions and a specific distance value, and the average or range of each feature used. Values of the features calculated for different distance values, *d*, and all four angular directions, could also be included in these average and range calculations.

3.3 Classification Algorithms

Statistical pattern recognition, including classification algorithms, plays an important role in the automated interpretation of images when the objects of interest are not easily separable from the image background or from other objects, and where additional complicating factors such as noise, occlusions etc. are present. Certainly the interpretation of the lung tissue samples used in this project is a challenging problem from the image processing perspective. For this project, two pattern recognition techniques are used: k-means clustering and discriminant analysis.

3.3.1 K-means Clustering

K-means clustering is a type of classification algorithm in which the goal is to categorize input data into clusters, such that:

- Each object is part of one and only one of k clusters
- Objects in the same cluster are as similar to one another as possible (based on some optimality criterion)
- Objects in different clusters are as dissimilar as possible

Consequently, k-means clustering requires some quantitative definition of "similarity". Where objects are to be classified based on a list of features, one similarity measure which is often used is Euclidean distance in multi-dimensional feature space.

An iterative form of k-means clustering is perhaps the most common [4]. In the iterative algorithm, the steps for classification of a set of input objects are as follows:

- An appropriate number of clusters, k, is selected. The centre of each cluster in feature space is determined.
- Each object is assigned to the closest cluster, based on distance to the cluster centre in feature space.
- Cluster centres are re-calculated based on the objects that have been assigned to the clusters.
- If any objects are now closer to a different cluster centre than that of the cluster to which they are currently assigned, these objects are re-categorized.
- The process of calculating cluster centres and re-assigning objects to clusters is repeated provided that any object has been moved to a different category during the iteration.

- Iterations continue until no objects are moved, or termination may also be triggered when the quality of clusters ceases to improve. This requires that a quantitative optimality criterion is defined to assess cluster quality.

There are a number of potential issues with this iterative algorithm, including the fact that it may not be apparent when iterations should stop. In fact, the algorithm may begin to oscillate between solutions which are suboptimal, but of equal quality. If a large number of iterations are required to arrive at the best solution, the algorithm can also become prohibitively computationally intensive.

Non-iterative clustering algorithms provide a less computationally intensive alternative. Of the non-iterative options, MacQueen's k-means algorithm is one of the best known [4]. In this algorithm, only two passes of the data are required. The first pass is used to locate the cluster centres, and the second pass is used to classify the data. The initial cluster centres used for the first pass through the data can either be chosen arbitrarily or by some more sophisticated technique. The sequence of steps of this algorithm is as follows:

- An appropriate number of clusters, k, is selected. The centre of each cluster in feature space is determined.
- First pass:
    - Each object is assigned to the closest cluster, based on distance to the cluster centre in feature space.
    - The centre of the relevant cluster is recalculated based on all points assigned to this cluster so far, whenever a new object is assigned to it.
- Second pass:
    - All objects are re-assigned, based on the minimum distance to the final cluster centres obtained in the first pass.

For this thesis project, MacQueen's k-means clustering algorithm is implemented to classify image pixels based on their RGB values, as one of the main steps in the detection of alveoli in the images. Refer to section 4.3.1 for details.

3.3.2 Discriminant Analysis

In discriminant analysis, objects are classified into mutually exclusive groups based on information from various predictor variables. A discriminant function is created from the

predictor variables, and may take any of a number of forms, including linear, quadratic, and Mahalanobis [15]. A linear discriminant function, for example, is constructed as the weighted sum of the individual predictor variables, as shown in equation 3.9.

$$L = w_1 x_1 + w_2 x_2 + \ldots + w_k x_k \tag{3.9}$$

where w_i = weights associated with each predictor variable
x_i = predictor variables

The discriminant function describes objects from the various groups on a one-dimensional scale. The discriminant score for each object to be classified is calculated by substituting the particular predictor variable values for that object. A cutoff score is then used to decide which group each object belongs to. Selection of appropriate values for the weights and cutoff score of the discriminant function is key, and needs to be done in such a way that the amount of overlap between the groups is minimized. The larger the difference in predictor variable values between the groups, the fewer classification errors will occur [16].

Linear discriminant analysis, as implemented in Matlab's *classify* function, was one of the methods used to distinguish areas of fibrin from areas of necrosis. Refer to section 4.3.3.3 for details.

3.4 Application of Image Processing Techniques to Histology Images

Application of image processing and computer vision techniques to biological images has had varying levels of success, and of application, with different medical imaging modalities. Within the histopathology field, a number of applications of image processing techniques are cited, including the following:

- Wurflinger et al. describe the use of eight image measurements, including eccentricity and bending energy, to distinguish between various types of cells [17].
- Both Tanaka et al. [18] and Pavlova et al. [19] discuss the applicability of conversion of histology images to the HSV colour model to interpret the images. Both groups of authors state that the advantage of the use of this colour model is that any deviations in processing samples with histology stains are isolated in the saturation and value components, while hue remains constant.
- Sutherland and Ironside [20] apply morphological operations to isolate areas of granular cells within histology images.

- The benefits of pairing a colour watershed technique with colour pixel classification schemes in order to improve results in classifying cellular structures in cytopathology images are examined by Lezoray and Cardot [21] in their paper.

- Image analysis work by Sutherland et al. [22] uses thresholding and morphological operations to locate and quantify features of Creutzfeldt-Jakob disease in brain tissue samples.

- Poulsen and Pedron [23] expand on their previous published work by experimenting with image analysis and pattern recognition methods that eliminate the need to explicitly segment images prior to measuring parameters and classifying objects. In both their earlier and more recent work, automated analysis of pap smear data mimics the manual analysis process by first locating the regions of interest in reduced resolution images, then performing detailed analysis of the selected regions at a higher resolution. Their previous work then segmented the high resolution images into background, cells and cell nuclei, followed by feature measurement and object classification. In their more recent work, the authors explore image processing techniques which do not require explicit image segmentation as an initial step.

- This classical sequence of image processing steps (image segmentation, feature extraction and object classification) is also applied by Liu et al. [5] in their examination of bone cross sections. As is also the case with the thesis project, the output of Liu et al.'s work is a set of segmented images, and quantitative measurements of features extracted from the images.

The goals of applying image processing techniques in the histopathology field range from cell counting and cell type classification, to deriving quantitative measurements of disease features from images, and automatically determining whether a disease is present within analyzed samples. Commercial products which apply image processing technology to automatically assess histologic and cytologic samples exist in a few areas where extremely high volumes are processed, such as analysis of blood samples, for example. An extensive literature search has revealed only one paper which discusses the application of image processing techniques to detect pulmonary disease. In this case, Sutton et al. [24] used texture properties to detect lung abnormalities in chest radiographs. No existing image analysis products or research has been found which specifically assess lung tissue samples, however.

CHAPTER 4
DATA AND METHODOLOGY

4.1 Data

Source data for this project consisted of cattle lung tissue samples on standard glass microscope slides (Figure 4.1). The tissue sample slides were selected from a slide collection owned jointly by the Western College of Veterinary Medicine (WCVM) and Prairie Diagnostic Services. Tissue samples used in this analysis were examples of cases of various pulmonary diseases, including several types of pneumonia and bronchitis.

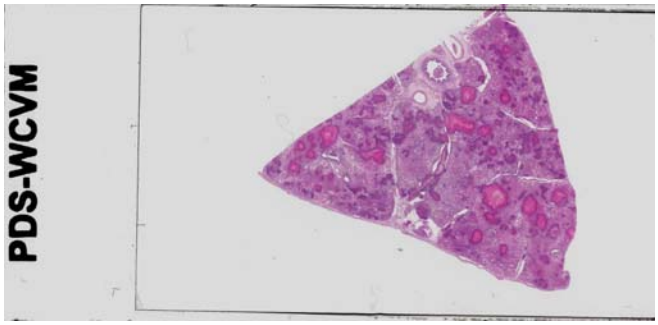


Figure 4.1: Cattle lung tissue sample slide

The tissue samples used in this project were prepared using hematoxylin and eosin stain (H&E), a standard histology stain. Hematoxylin stains basophilic structures, such as cell nuclei, a deep purple colour. Eosin is an acidic stain with a red colour. Eosin stains acidophilic structures, such as red blood cells, collagen fibres etc., a pink or red colour. However, use of this or any other pathology stain can result in colour intensities which are highly variable. The same cellular structures can appear a red colour on one slide, and pink on another, for example. Although colour was an important cue in the automated analysis of the tissue samples, only relative, not absolute, colours could be used due to these variations.

A total of 10 tissue sample slides were included in the evaluation stage of this project. This set of tissue sample slides was selected by Dr. G. Appleyard and Dr. G. Wobeser, two experts in Veterinary Pathology at WCVM.

Table 4.1: Slides analyzed as part of this project

Slide Identifier	**Pathologist's diagnosis of disease(s) present in sample**
D03-10538	Necrotizing pneumonia; fibrinous pleuritis
D96-4925	Fibrinopurulent pneumonia; interstitial pneumonia
D03-00716	Chronic necrosuppurative bronchopneumonia
D03-11282	Suppurative and necrotizing bronchopneumonia
D03-45740	Severe bronchointerstitial pneumonia with extensive multifocal bacterial necrosis and bronchiectasis
D03-36878	Fibrinous, hemorrhagic and necrotizing pleuropneumonia
D03-51213	Bronchiolitis obliterans
D03-37303	Bronchopneumonia, locally extensive with coagulation necrosis and sequestrum
D03-51216	Acute necrotizing and fibrinous pneumonia and bronchiolitis with early Mycoplasma bovis lesions
N95-4825	Bronchiolitis obliterans; Bovine respiratory syncytial virus

Digital images of all tissue samples were created using a resolution of approximately 0.5 μm per pixel (refer to details in Section 4.2.1). Resulting images ranged in size from approximately 24,000 by 26,000 pixels up to approximately 46,000 by 62,000 pixels. Due to memory limitations, processing of "subimages" was required rather than processing entire tissue sample images at once. The images were processed in sections of 2000 by 2000 pixels. For the final analysis of this project, ten 2000x2000 pixel subimages were randomly selected from each of the tissue sample images. Although the digital images include only the least bounding box of the tissue of interest, since tissue samples are irregularly shaped and the generated digital images are rectangular, subimages selected from around the boundaries of the image may not contain any portion of the tissue sample. All blank subimages were eliminated, and additional subimages were randomly chosen until a set of 10 usable subimages per tissue sample was created. From these sets of images, Dr. G. Wobeser then selected, for each tissue sample, 5 images which were suitable for analysis. Images were rejected at this stage if the image area predominately consisted of tissue other than lung parenchyma, areas which are not normally included in the manual analysis of the disease state of the lung.

For the automated image analysis, determining whether any areas of the image corresponded to regions of the slide outside the tissue sample was a necessary preprocessing step. These background areas surrounding the tissue sample should be excluded from processing (Figure 4.2). Background areas in the images will be close to white in colour. However, it is

important that any air-filled alveoli (which will also be "white" regions) which happen to lie along the borders of the subimages not be misinterpreted as background regions. These alveolar regions should be included in the tissue area within the image. As a simple method for distinguishing true background regions from partial alveoli, the following steps were implemented. First, the image foreground region, corresponding to the tissue, was identified and its boundaries were found. Next, each of the four borders of the image was examined in turn. If any region boundary pixels occurred along that border, the border was marked for further processing. For each of these borders, an additional row (or column) of non-white pixels was temporarily added to that edge of the image. These additional pixels would then form the missing boundary of any truncated alveoli lying along that border, enclosing these regions as "holes". A morphological fill operation (as implemented in Matlab's *imfill* function) was used to fill all holes in the foreground region, and then the extra row or column of pixels was removed. This process was repeated for each of the borders which contained region boundary pixels. Finally, an image was created indicating all pixels belonging to the foreground region, which would now correspond to the full tissue contents of the image including truncated alveoli lying along the image borders. This labeled image was then used as input to all subsequent processing steps applied to the image, in order that each stage of the analysis could operate exclusively on the areas of the image which corresponded to the tissue, and exclude areas which corresponded to background regions of the slide. This simplistic technique for determining background areas has obvious shortcomings, including the fact that a truncated alveolus which happens to lie at the corner of an image would be erroneously excluded as part of the background (Figure 4.3). However, any inaccuracies introduced by this simple background determination technique were judged to be minor compared to the cost of implementing a more sophisticated method for distinguishing background areas from tissue.

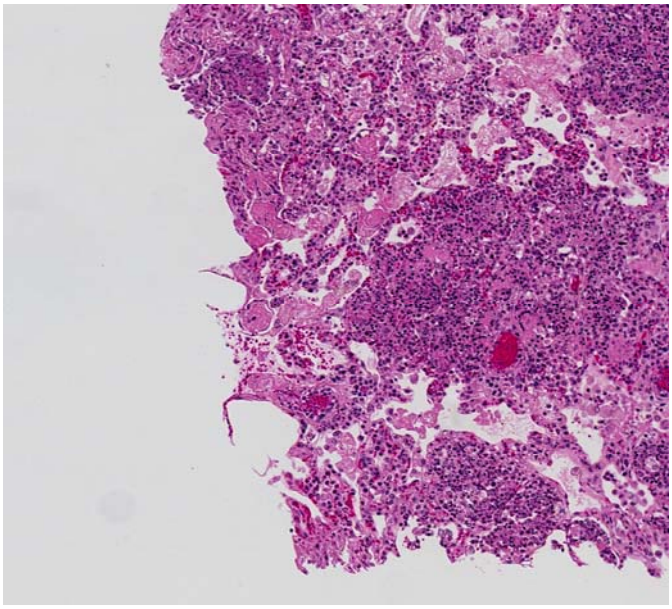


Figure 4.2: Example of a 2000x2000 pixel subimage with background areas which will be excluded from further processing

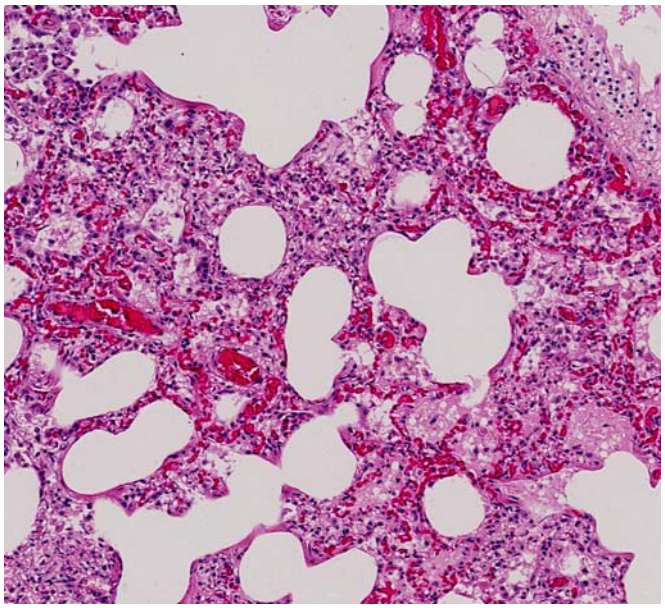


Figure 4.3: Example of a 2000x2000 pixel subimage with alveolar regions along the border of the image, which will be included in further processing

Throughout this paper, images will be referred to using the row and column co-ordinates (see section 3.2.1) of the upper left corner of the corresponding region within the overall slide. For example, image [12001, 20001] of slide D03-10538 refers to the 2000x2000 pixel block which contains pixels from rows 12001 to 14000, and columns 20001 to 22000.

4.2 Materials

4.2.1 Digital Slide Scanner

Digital images were created from the tissue sample slides using a Trestle Corporation MedMicro system (Figure 4.4), including the company's Digital Slide Module software. The MedMicro system includes a CCD camera mounted on a microscope assembly, a robotic slidebed, multiple microscope objectives, a light source, computer and MedMicroscopy Viewer software. In the configuration used for this project, the objectives included were 2X, 4X and 10X Universal Plan Fluorite Objectives. The scanner's robotic slide bed moves the slide beneath the microscope during image capture, allowing multiple field-of-view images to be obtained. Trestle's scanner software then seamlessly stitches these multiple images together to create a single image of the entire tissue sample. The ability to capture a high resolution single image of an entire tissue sample is a unique feature of this equipment, which is not available with most digital slide imaging products.

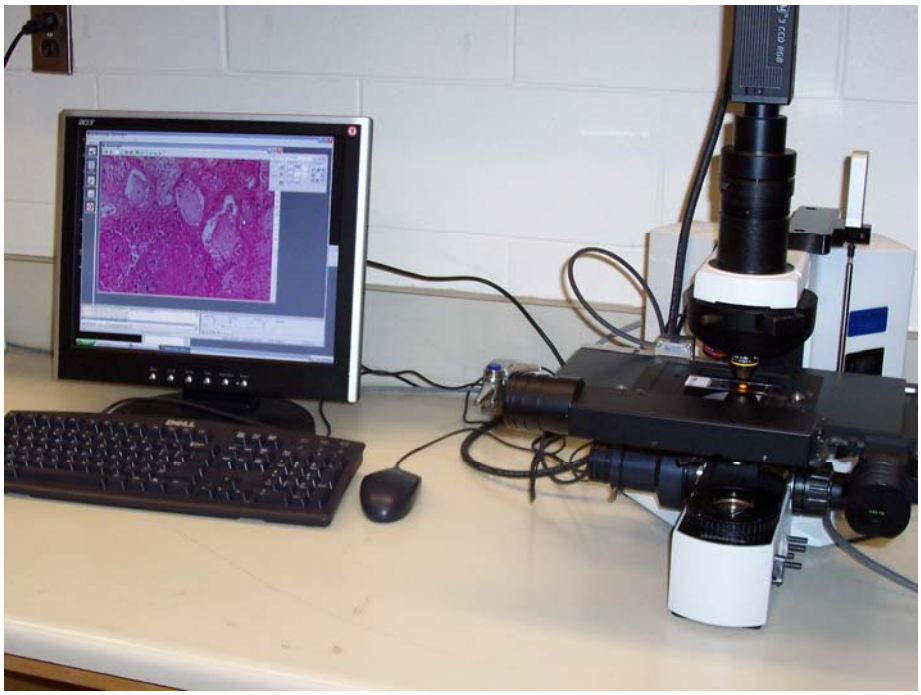


Figure 4.4: Trestle MedMicro system

The MedMicroscopy Viewer software controls operation of the scanner equipment, as well as providing the user with the ability to view and navigate images either locally or remotely.

Selection of the microscope objective, focal plane selection, focusing, navigation and annotation of images is all controlled through this software.

The Digital Slide Module software is integrated with MedMicroscopy Viewer. The Digital Slide Module software captures and stores digital images of the slide contents. The images created use 24-bit colour, and are stored as TIFF files using lossless compression.

The MedMicro equipment configuration used for this project is capable of creating images with resolution of up to approximately 0.5 μm per pixel. Cells of the CCD camera capture a distance of 4.65 μm per pixel, without magnification. Using the 10X objective, each pixel dimension corresponds to a distance of 0.465 μm on the tissue sample. For this project, all images were created using the 10X objective.

4.2.2 Programming Language

All programs for this project were written using Matlab (www.mathworks.com) version 7.0.4.352, Service Pack 2, including Matlab's Image Processing Toolkit, version 5.0.2.

4.3 Methodology

Three main branches of processing were used to detect the subset of disease indicators, as summarized in Figure 4.5. These three techniques are described in sections 4.3.1, 4.3.2 and 4.3.3.

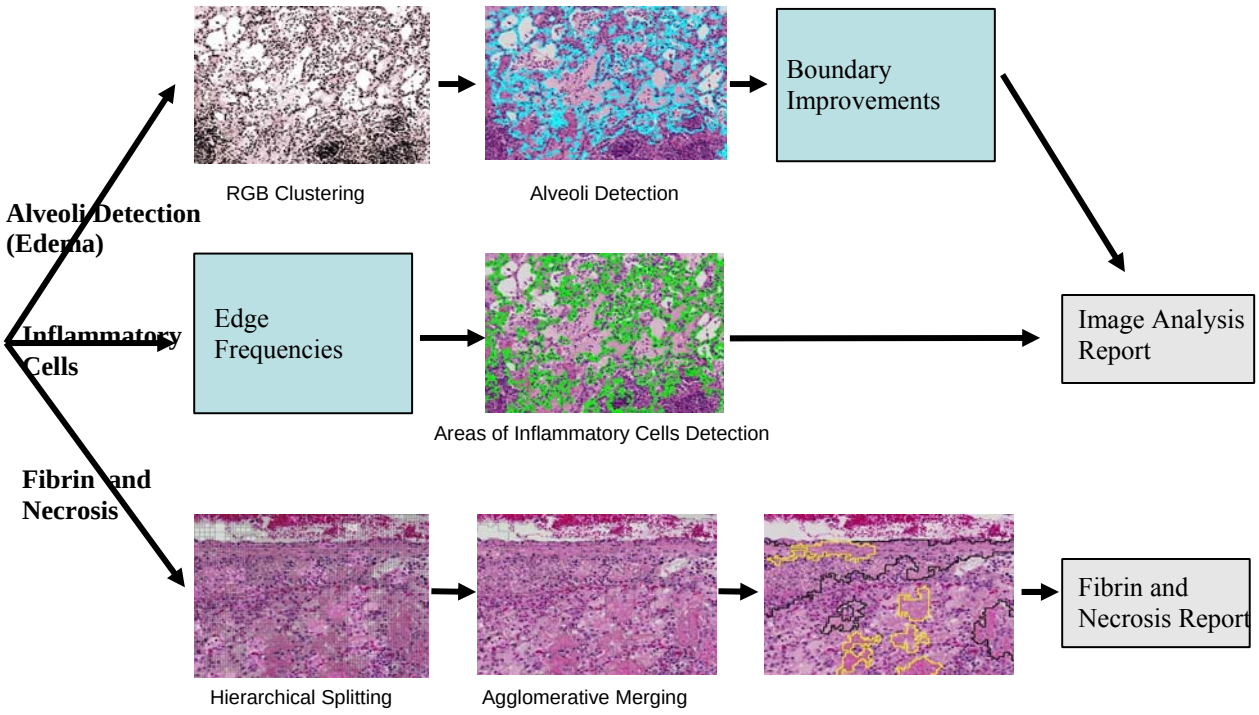


Figure 4.5: Summary of processing used to locate disease indicators

4.3.1 Detection of Alveoli and Alveolar Contents

4.3.1.1 Determining Alveolar Boundaries

In a normal lung, all alveoli would be air-filled resulting in their interiors appearing white in the corresponding tissue sample image. In diseased lungs, alveoli may be air-filled or may be partially or completely filled with fluid. With the standard H&E staining used to prepare these slides, fluid will appear a light pink to darker shade of pink depending on the amount of protein content. Alveoli in the diseased samples may also have some infiltration of neutrophils, resulting in dark purple nuclei of these cells appearing in the alveolar spaces in the tissue samples.

As a first step in detecting alveoli, a clustering algorithm was applied to all image pixels based on their RGB values. MacQueen's k-means clustering algorithm was used (see section 3.3.1). The intention of this clustering was to categorize each pixel as belonging to one of three categories, as shown in Table 4.2.

Table 4.2 Clusters based on RGB values

Cluster Name	**Image colours to be included**	**Corresponding tissue content**
White	White, light shades of pink	Air, fluid
Pink	Darker shades of pink	Fibrin, necrosis, alveolar walls
Purple	Dark purple	Nuclei of neutrophils, macrophages

Euclidean distance in RGB space was the distance metric used to determine the appropriate cluster for each pixel, although for computational efficiency, the square of the Euclidean distance was computed. The end result of this pixel classification should be three clusters within RGB space where image pixels within each cluster are as similar as possible to one another, and as dissimilar as possible to pixels in any other cluster. By grouping all pixels corresponding to air and fluid together, the intention is to separate out regions corresponding to inner alveolar space.

Starting values for the clusters' centres, or exemplars, were calculated by sampling a number of pixels belonging to each category in one sample image, and taking the average red, green and blue values for these sample points. The starting exemplar values used were as follows:

White cluster	(225, 225, 225)
Pink cluster	(210, 145, 180)
Purple cluster	(120, 90, 160)

It is important to note that these starting exemplars are not necessarily good representative points for the set of images to be analyzed. In fact, due to staining variations within tissue samples, these points may not even be appropriate representative points for all regions of the image from which they were calculated. However, appropriateness of the starting exemplar values is not particularly relevant. As mentioned in the description of MacQueen's k-means clustering included in section 3.3.1, starting exemplar values could be chosen randomly. Although the starting values are not necessarily appropriate for any particular slide being analyzed, at the end of the first pass of this algorithm the final exemplar values are representative of the range of pixel values assigned to each cluster from that particular image. These image-appropriate exemplar values are then used to re-classify all pixels in that image in the second pass of the algorithm. Since staining variations between the tissue sample slides was a known challenge in the analysis of images of this project, the sample-specific selection of cluster

exemplar values included in k-means clustering made this type of classification an appealing choice. Implementation of MacQueen's k-means clustering algorithm in particular was selected based on ease of implementation, and efficiency of processing.

At the end of the RGB clustering step, each pixel had been assigned to one of three clusters. A new image was created in which each pixel was assigned a colour (white, pink or black) representing the cluster to which it belonged (Figure 4.7).

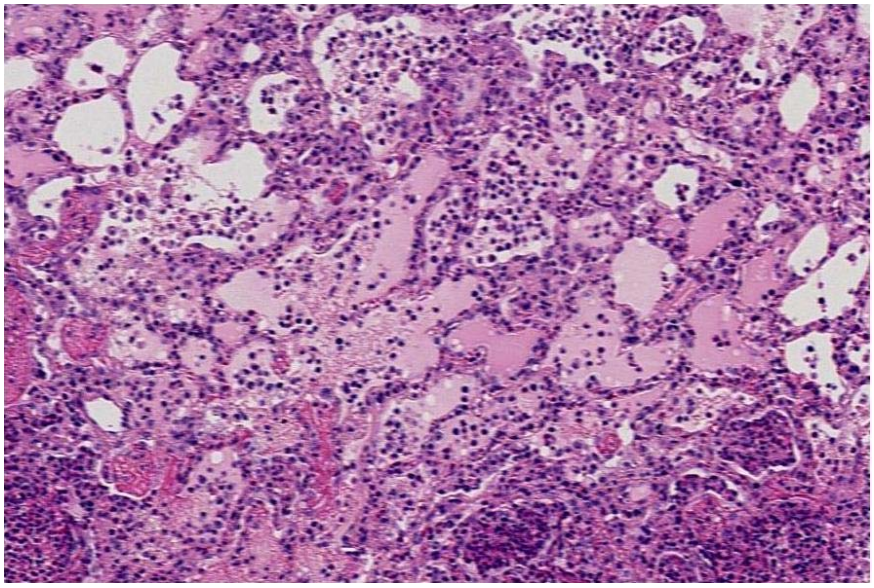


Figure 4.6: Sample image from one of the diseased lung tissue samples

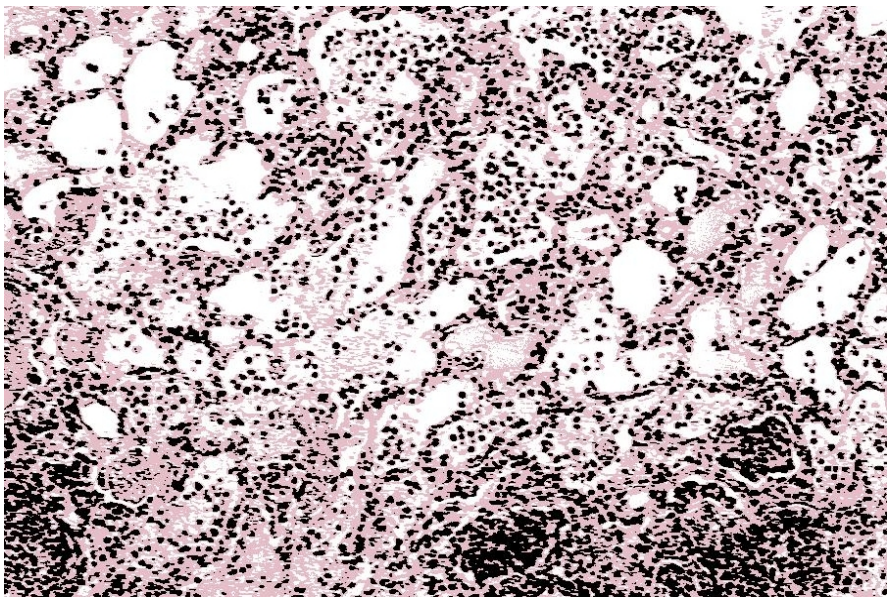


Figure 4.7: RGB clustered image created from previous image. Pixels assigned to the white cluster are shown as white pixels; pixels assigned to the pink cluster are shown as pink; pixels assigned to the purple cluster are shown as black.

In the RGB clustered image, all pixels which correspond to either air or fluid in the tissue sample should be shown as white, meaning that inner alveolar space in this new image should be primarily white. The RGB clustered image was then used as the input image in detection of alveolar boundaries. Thresholding was applied to the RGB clustered image to create a binary image in which all white pixels remained white, and all pink and black pixels were converted to black. Since in binary images objects are typically displayed in white against a black background, in this binary image alveoli are the objects and all else is background. A fill operation (Matlab's *imfill*) was applied to remove holes in image regions, and a morphological opening operation (Matlab's *bwareaopen*) was used to eliminate any alveolar regions which consisted of less than a minimum number of pixels. Very small regions in this image likely correspond to collapsed alveoli within the tissue sample. These small alveoli are not assessed as part of the tissue analysis, and so were removed in order to reduce clutter in the final alveolar boundary image. Two choices for the minimum significant size of alveoli to be detected were set: 200 pixels for high-level (more precise) detection, and 1000 for lower-level detection. The 200 pixel minimum was chosen arbitrarily. The 1000 pixel limit was selected in consultation with Dr. G. Wobeser on the minimum alveolus size which would be marked as part of the manual image analysis.

Regions within the binary image were then labeled using Matlab's *bwlabel* routine, a connected components labeling algorithm [11]. Use of 4-connectivity with the connected component labeling routine produced preferable results over use of 8-connectivity. When 8-connectivity was used, so that all 8-connected neighbouring foreground pixels of any region pixel were also considered part of that region, results showed that many alveoli were combined into larger regions. With 4-connectivity, the number of individual alveoli which were combined was reduced. As a first attempt at locating the alveolar boundaries, a boundary tracing algorithm (Matlab's *bwboundaries*) was applied to the labeled image (Figure 4.8).

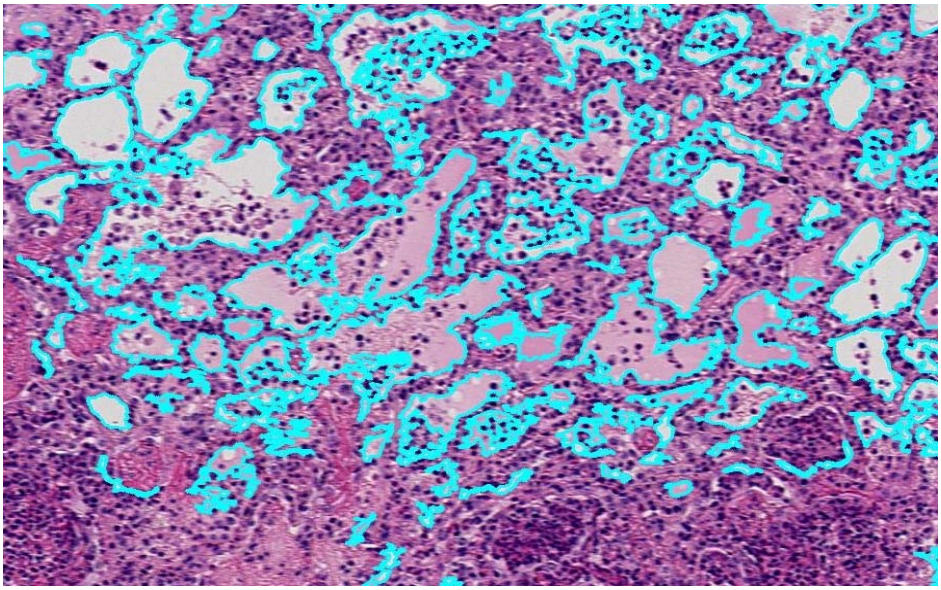


Figure 4.8: First attempt at alveolar boundary detection

Results of this preliminary alveolar boundary detection were encouraging, although some obvious problem areas existed. The most obvious problem occurred where alveoli had been infiltrated by neutrophils. In these cases, if the nuclei of the neutrophils were connected to the edge of the alveolus in the image, the nuclei were interpreted as belonging to the alveolar walls instead of being inside the alveolus. As a result, inaccurate alveolar boundaries were traced in a number of areas, as illustrated in Figure 4.9.

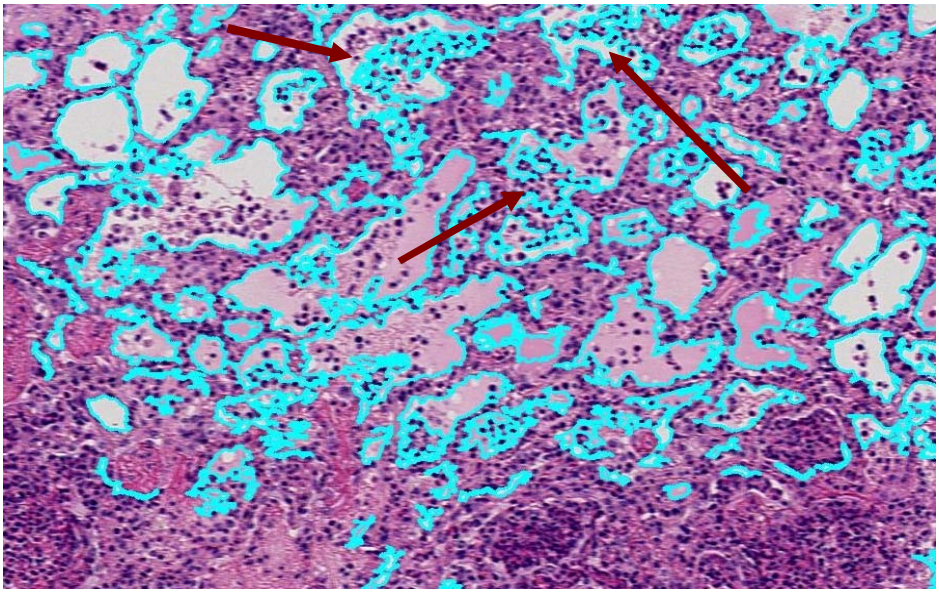


Figure 4.9: Red arrows mark areas where alveolar boundary detection is inaccurate due to neutrophils inside alveoli

Elimination of these boundary inaccuracies is problematic. Unfortunately, nuclei inside alveoli cannot just be located and temporarily removed, since at this point there is no accurate definition of inner alveolar space. These nuclei are currently interpreted as part of the alveolar walls, not as interior to the alveoli. It is also not possible to subtract all nuclei from the image, since this would result in large areas of inflammatory cells being interpreted incorrectly as alveoli. A distinguishing characteristic of the nuclei of neutrophils inside alveoli from those in regions of inflammation is that inside the alveoli the nuclei tend to be separated from one another. In areas of inflammation, tight clumping of these nuclei tends to occur. Although this is not always the case, this fact was used to temporarily remove the nuclei from within alveoli to improve boundary detection. The following sequence of steps was used, as summarized in Figure 4.10.

The area of individual nuclei within the images varies, but was approximated as 250 pixels. A second binary image was created from the RGB clustered image, by converting all white and pink pixels to white, and leaving all black pixels as black. The inverse of this image was then taken, resulting in an image (the inverse image) in which only nuclei pixels were white and all else was black. A morphological opening operation (as implemented in Matlab's *bwareaopen* function) was then applied to this image to remove regions which were less than 250 pixels in size, in order to remove isolated nuclei. The resulting image then had only clumps of nuclei indicated in white; all else was black. This opened image was subtracted from the inverse image, producing a difference image. Since pixels corresponding to clumps of nuclei were white in both images, these pixels ended up being black in the difference image ((255,255,255) – (255,255,255) = (0,0,0)). Pixels which were originally white or pink in the RGB clustered image were black in both the inverse image and the opened image, and so ended up black in the difference image ((0,0,0) – (0,0,0) = (0,0,0)). Pixels corresponding to individual nuclei were white in the inverse image and black in the opened image. As a result, isolated nuclei ended up being white in the difference image ((255,255,255) – (0,0,0) = (255,255,255)). The difference image thus indicated all isolated nuclei which were to be removed, since only these nuclei were white, all else was black. A dilation operation was applied to the difference image to slightly enlarge all nuclei within this image, and the resulting image was then added to a new binary image, in which all white pixels from the RGB clustered image remained white, but all pink and black pixels were converted to black. Adding black pixels (RGB value (0,0,0)) to any pixel in

the binary image does not change its value. Thus the addition of the difference image only affected the pixels of the isolated nuclei. In the original binary image, these pixels were black; in the difference image these pixels were white. As a result, in the sum image the pixels of isolated nuclei became white ((0,0,0) + (255,255,255) = (255,255,255)). These isolated nuclei within alveoli were thus effectively removed from this binary image. This modified binary image was then used as the input image for the connected components labeling and boundary tracing steps.

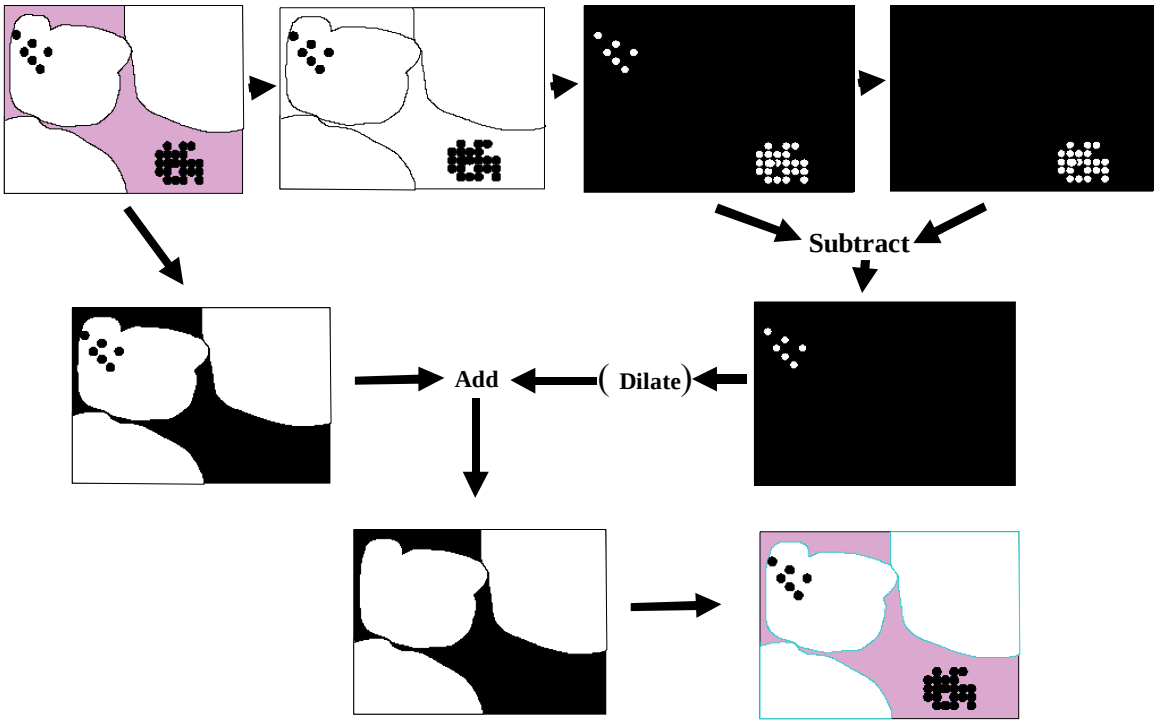


Figure 4.10: Additional processing steps to improve alveolar boundary detection

Although this processing does not eliminate all problems with boundary detection caused by nuclei inside alveoli, boundary tracing results were greatly improved in some areas (Figure 4.11).

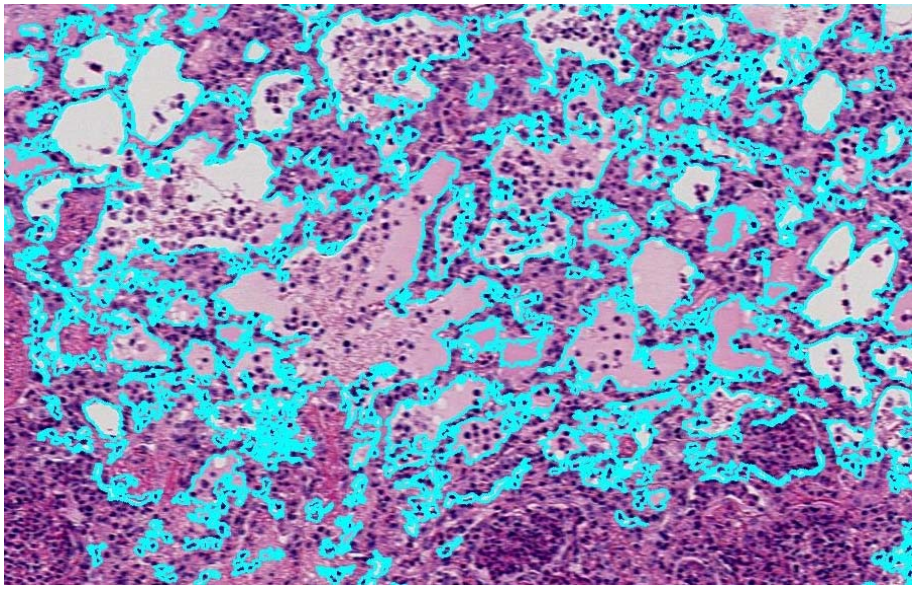


Figure 4.11: Improved alveolar boundary detection

One additional attempt was made to further improve these results. Although whole areas of nuclei within alveoli are no longer interpreted as part of the alveolar walls, clumps of nuclei which are directly adjacent to true alveolar walls are still, in some cases, interpreted as part of those walls. In order to break the small "threads" that bind these interior nuclei to the outer edge of alveoli, a 3x3 averaging filter

$$1/9 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

was applied to the sum image after the isolated nuclei were removed. The filtered image was then converted to binary, and used as input to the connected components labeling and boundary tracing steps described above. Results of adding this step were reasonably good. Some problem areas within alveoli were eliminated, resulting in more accurate alveolar boundary detection. However, the disadvantage of using this step was that in some cases alveolar regions were combined inaccurately into a single large alveolus.

4.3.1.2 Assessment of Alveolar Contents

In the assessment of disease conditions present in the lung tissue samples, edema is a key indicator. Once the alveolar regions have been determined, as described in the previous section, the pixels within these regions can be evaluated to determine whether the alveoli are air-filled or fluid-filled, and to calculate the amount of fluid present. This assessment proceeds as follows.

The connected components labeling step produced a labeled image in which all pixels belonging to each alveolus are indicated. Each labeled region (alveolus) was then examined separately. Any pixels within the region which had been categorized as part of the "Purple" cluster in the RGB clustered version of the image were excluded from processing. All remaining pixels within the alveolus were then re-categorized as being "white" or "pink" using new exemplar values. This re-classification was necessary since in the first clustering pass, pixels corresponding to both fluid and air were grouped together as part of the white cluster. At this point, the essential distinction is between pixels corresponding to air, and those corresponding to fluid. Exemplar values used for the new white and pink clusters were:

White cluster	(245, 245, 245)
Pink cluster	(235, 195, 215)

Once all pixels within the alveolus were classified, the percentage of the total pixels within the alveolar region which were considered pink (i.e. fluid) was calculated. Any alveolus which contained at least 33% pink pixels was considered significantly fluid-filled. These fluid-filled alveoli were indicated in a separate image, created as part of the output of the software (Figure 4.12). The project software also produces a report indicating the percentage of alveoli which are fluid-filled, and the percentage of total alveolar space which is fluid-filled.

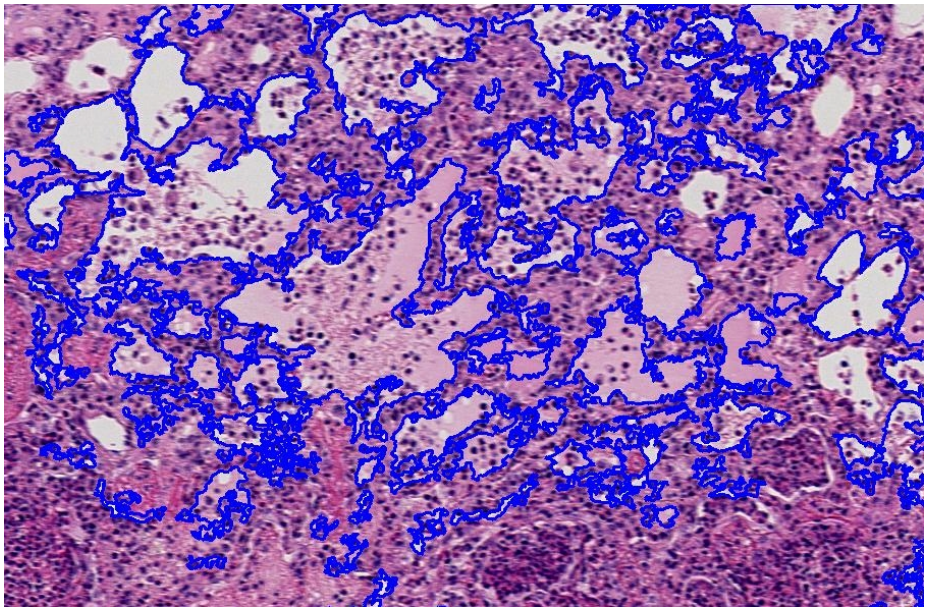


Figure 4.12: Alveoli with 33% or greater fluid content

4.3.2 Detection of Areas of Inflammatory Cells

Areas of inflammation in the lung tissue samples are characterized by the presence of neutrophils and/or macrophages. Although some number of these cells will always be present in the lung, a high concentration of these inflammatory cells in any area is indicative of disease. The detectable feature of this disease indicator is thus clumps of nuclei of these leukocytes. Image regions which contain large numbers of nuclei will have a high number of edges within a small area. Consequently, texture properties, including edge frequency and edge density, were used to find these image areas.

Edge detection in colour images is much more complicated than in greyscale images. As a simplifying measure, for this analysis tissue sample images were converted to greyscale using Matlab's *rgb2gray* function, prior to locating edges.

To find the edge frequency image, the first step was to calculate the edge magnitude for each pixel in the image. Pixels where the edge magnitude exceeded a certain threshold were marked as foreground pixels in this image; all other pixels were set to background. Edge frequency images reveal properties of the underlying texture in that the finer the texture, the larger the number of edges within a given area.

To assess the edge magnitudes at each pixel, a distance-dependent gradient function, *g(d)*, was used. For image *f* and distance *d* this texture description function is defined as [11]:

$$g(d) = \left|f(i,j) - f(i+d,j)\right| + \left|f(i,j) - f(i-d,j)\right| + \left|f(i,j) - f(i,j+d)\right| + \left|f(i,j) - f(i,j-d)\right| \quad (4.1)$$

The gradient function was calculated at all image pixels, using a distance of 6. This particular distance was selected based on the fact that nuclei within the images typically have diameters of approximately 12 to 13 pixels. Using a distance of 6 ensures that, at a minimum, pixels near the centres of nuclei will have high gradient values. Pixels towards the boundaries of nuclei may or may not have high gradient values, depending on the proximity of the particular nucleus to its neighbouring nuclei.

Evaluation of the gradient function on sample images revealed that maximum gradient values were typically 800 to 1000. An appropriate threshold value of 250 was determined experimentally, and the edge frequency image was created by marking all pixels with a gradient value greater than or equal to this threshold. The next step in this processing was to determine the number of edges within each small neighbourhood within the image, to confirm that the detected edge pixels occurred in areas of high edge frequency. For this evaluation, 3x3 neighbourhoods were used. For each pixel, if the 3x3 neighbourhood centered at that pixel contained at least 6 significant edge pixels in the edge frequency image, that pixel was considered part of a dense edge area and marked as a foreground pixel in the edge density image. Otherwise, the pixel was labeled as background. The resulting image was then assessed to find regions of high edge density. These regions, it was hoped, would correspond to areas of inflammatory cells within the tissue sample images. A morphological fill operation (as implemented in Matlab's *imfill* function) was used to fill holes in the regions, and small regions were eliminated using a morphological opening operation (as implemented in Matlab's *bwareaopen* function). For this assessment, one of two different minimum region sizes could be specified. For high-level, more precise, region detection, areas of fewer than 1000 pixels were eliminated. For a lower level of detection, areas of fewer than 2500 pixels were removed. Both minimum size threshold values were selected based on examination of areas of inflammatory cells in sample images. Regions were then labeled using a connected components labeling routine (Matlab's *bwlabel* function) with 8-connectivity specified, and region boundaries were traced using Matlab's *bwboundaries* function.

As a confirmation that the high edge density regions detected by the above processing actually corresponded to areas of inflammatory cells within the tissue samples, a further check was added. For each pixel within a high edge density area in the edge density image, its corresponding pixel in the RGB clustered image produced as part of the alveoli detection step was found. The percentage of pixels within each high edge density region which were classified as part of the purple cluster in the RGB clustering step was evaluated. Through analysis of a series of sample areas of inflammatory cells, it was determined that these areas typically contain at least 33% "purple" pixels. Any high edge density areas found during this processing which had less than this percentage of purple pixels was considered inaccurately identified. Although this check was added to the areas of inflammatory cells processing, it was found that this condition was rarely true. In a test of one sample image in which 134 high edge density regions were identified, only one of these regions contained less than this percentage of nuclei pixels, for example.

The final evaluation of the success of this technique in locating areas of inflammatory cells was to superimpose the detected boundaries of these areas onto the original image (Figure 4.13). These highlighted images revealed that the detected areas did indeed correspond to areas containing high numbers of nuclei which are typical of inflamed tissue regions.

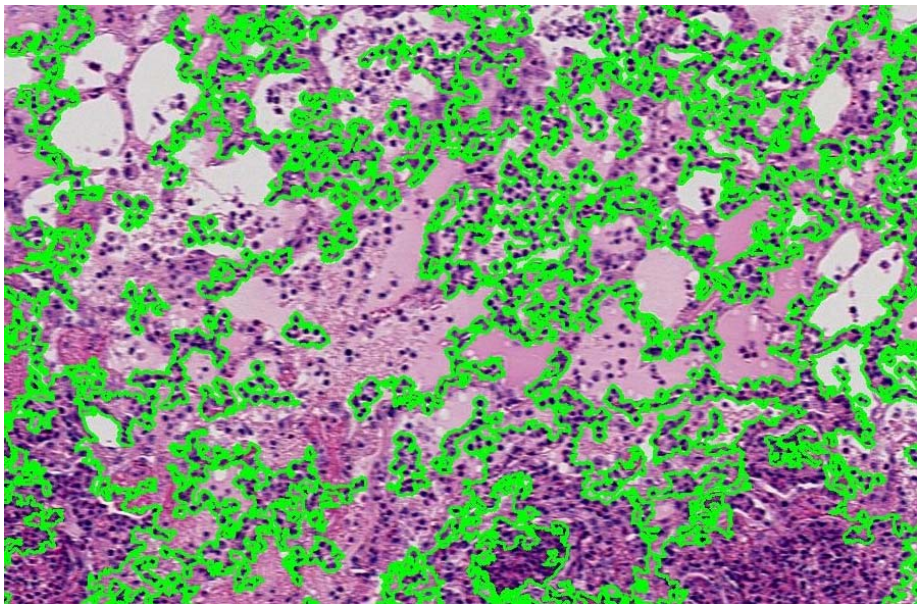


Figure 4.13: Detected areas of inflammatory cells

The project software calculates the total image area occupied by the regions that contain inflammatory cells. The quantitative information produced by the software for this part of the analysis is the percentage of overall tissue area covered by these regions.

4.3.3 Detection of Areas of Fibrin and Necrosis

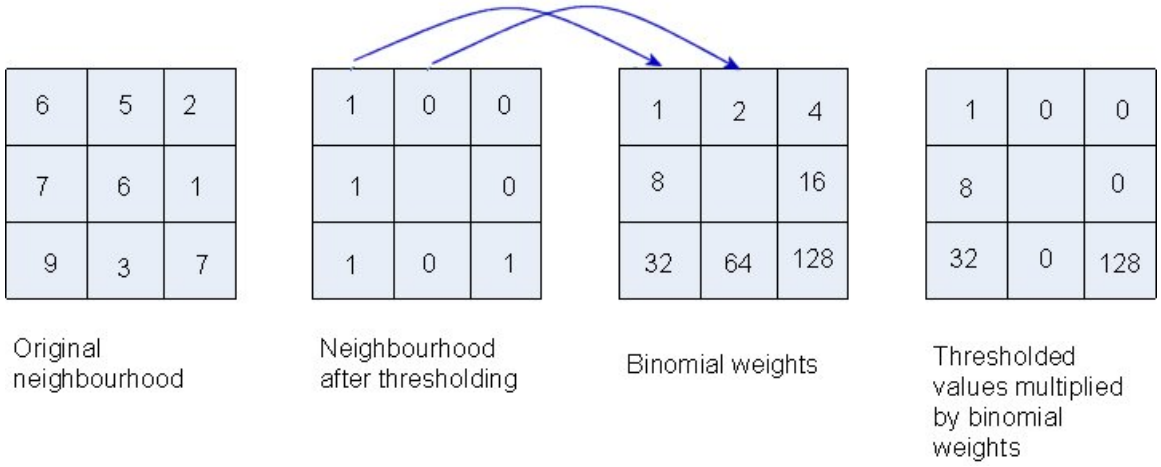
4.3.3.1 Ojala-Pietikainen, Chen-Chen method of segmentation based on colour and texture

Most image processing techniques for segmentation of images use either colour or texture independently. However, it has been shown that combined use of spectral and textural features can improve the accuracy of image region classification [3]. In their paper, Ojala and Pietikainen [25] proposed a method of unsupervised texture segmentation which uses distributions of local binary patterns as a texture description, and contrasts these patterns to distinguish between neighbouring regions during the segmentation process. Although the authors claimed that their technique could be easily generalized to use additional information, including colour features, their paper described the application of this technique to greyscale images only. Chen and Chen [26] extended the work of Ojala and Pietikainen to combine the use of colour and texture feature distributions to segment images. Both the Ojala-Pietikainen method, and the Chen-Chen extension to it, are unsupervised segmentation techniques, in that no a priori information about the number and types of textures, or the number of regions, is required.

Ojala-Pietikainen Method

In the image segmentation method described by Ojala and Pietikainen, the texture of an image region was defined by a combination of the local binary pattern (LBP) and the contrast (C). The local binary pattern was used to describe the spatial structure of the local texture. The LBP was computed at each pixel within an image region, using a 3x3 neighbourhood centered at that pixel. First, the 3x3 neighbourhood was thresholded using the value of the centre pixel. A binary result was assigned to each pixel in the neighbourhood, with 1 indicating that the original pixel value was greater than or equal to that of the centre pixel, and 0 indicating that the pixel value was lower. Next, the thresholded neighbourhood values were multiplied by binomial weights, and the resulting values were summed to obtain an LBP number for this

neighbourhood. A contrast number was calculated for the neighbourhood by computing the difference between the average grey-level value of all pixels in the neighbourhood which were thresholded to value 1, and the average grey-level value of the neighbourhood pixels which were thresholded to 0. The ratio LBP/C was assigned to the centre pixel of the neighbourhood.



LBP = 1 + 8 + 32 + 128 = 169
C = (6 + 7 + 9 + 7) / 4 – (5 + 2 + 1 + 3) / 4

Figure 4.14: Example of LBP and C calculations for a 3x3 neighbourhood, taken from Ojala, Pietikainen paper [25].

Once an LBP/C number had been calculated for each pixel within an image region, the LBP/C distribution for the region was summarized in a histogram. The size of the histogram would be 256 multiplied by the number of possible values for contrast, C. For their work, the authors (Ojala and Pietikainen) chose to limit the number of contrast values to 8.

As an obvious consequence of the use of the binomial weights matrix to calculate LBP values, the LBP/C descriptor is rotationally variant. This may not be appropriate for various images.

The LBP/C histogram was then used as the measure with which to compare image regions. The G statistic, a log-likelihood ratio [27], was used to compare the LBP/C distributions of image regions. The higher the value of this statistic, the less similar the two histograms were, and so the less alike the corresponding image regions were.

Comparisons of the LBP/C histograms of image regions formed the basis for image segmentation in the Ojala-Pietikainen method. The segmentation part of their algorithm involved three steps: hierarchical splitting, agglomerative merging, and pixelwise classification.

Hierarchical splitting was used to divide the image into regions of roughly uniform texture. Agglomerative merging then combined similar adjacent regions until a stopping criterion was met. Completion of these two steps yielded an estimate of the various textured regions within the image. Pixelwise classification was used as a final refinement, to improve the accuracy of the textured region boundaries determined in the previous steps.

In the hierarchical splitting step, the image was recursively split into blocks of decreasing size, until each block enclosed a region of fairly uniform texture. Initially the image was divided into blocks of a certain maximum size, which, through experimentation, the authors decided to set to 64 by 64 pixels. This initial partitioning was seen as a necessary step, since if the texture uniformity test were applied to very large image segments, it was possible that a large region containing numerous small texture patches evenly distributed in the region could be evaluated as uniform, and thus not be subdivided. Each of these initial image subregions was then evaluated to determine texture uniformity of that region. Each block was temporarily divided into four equal subblocks, and the LBP/C histogram was calculated for each of the subblocks. The G statistic was measured for each of the six pairings of the four subblocks, and the ratio of the largest and smallest values of the six G statistics was used as the homogeneity measure for the block (equation 4.2).

$$GRatio = G_{\max} / G_{\min} > G\_Threshold \tag{4.2}$$

If the maximum G statistic divided by the minimum G statistic was greater than a certain threshold, the block was divided into its four subblocks, and the homogeneity test was repeated with each of the subblocks in turn. If the ratio was less than the threshold value, the block was considered uniform in texture, and was not subdivided. For their work, the authors determined experimentally that a ratio of greater than 1.2 indicated texture inhomogeneity. Hierarchical splitting of the image continued until each remaining block had been determined to be homogenous, or it had reached the minimum block size, which the authors set to 16 by 16 pixels.

The authors did note that the hierarchical splitting step was not a mandatory one. The algorithm could also be implemented by dividing the image into blocks of the minimum block size (16 by 16), and then having the agglomerative merging step take care of combining neighbouring regions with similar texture properties. However, the authors did find that using

hierarchical splitting to locate larger areas of uniform texture improved the convergence of the agglomerative merging step.

In the agglomerative merging step, neighbouring regions with similar texture properties were combined until a stopping criterion was reached. Adjacent regions were merged in order of increasing merger importance (MI_{OP}), where merger importance for a pair of regions was defined as the G statistic for the two regions, multiplied by the number of pixels in the smaller of the two regions:

$$MI_{OP} = p \times G \tag{4.3}$$

where p = number of pixels in the smaller of the two regions

Merging regions in this order selected at each step the merger which would result in the least amount of change to the partitioned image. When two regions were merged, the LBP/C histograms of the regions were combined, and the G statistics for this newly merged region and each of its neighbouring regions were calculated, before deciding on the next merger to be performed. Merging proceeded until the ratio of the current best merger importance value to the largest merger importance value of any previous merger exceeded a specified threshold (equation 4.4).

$$MIRatio_{OP} = MI_{cur} / MI_{max} > MI\_Threshold \tag{4.4}$$

The authors determined experimentally that this threshold, *MI_Threshold*, should be set to 2.0.

On completion of the agglomerative merging step, the image was segmented into regions of roughly homogenous texture. Each region, however, had jagged boundaries as a result of combining block shaped image regions of various sizes. To improve the accuracy of region boundaries, pixelwise classification was used. In the first pass of this step, each pixel which was 4-connected to two or more texture regions (i.e. occurred on the boundary of two or more texture areas) was examined. A disc of a specified radius was placed on each of these pixels, and the LBP/C histogram computed for the disc. This histogram was then compared to the LBP/C histograms of all 4-connected texture regions. If the G statistic for the disc and one of the neighbouring texture regions to which the pixel did not belong was lower than the G statistic for the disc and the texture region to which the pixel currently belonged, the pixel was relabeled to

be part of this neighbouring texture region. The LBP/C histograms of both regions were updated accordingly. In the next pass, only the neighbourhoods of the pixels which were relabeled in the first pass were examined. Pixelwise classification continued until no pixels were relabeled in a pass, or until the maximum number of passes was reached. For their work, the authors set the maximum number of passes to be 32. Through experimentation, they determined that a disc radius of 11 pixels was optimal.

Chen-Chen Modification of the Ojala-Pietikainen Method

In the Ojala-Pietikainen method of image segmentation, distributions of texture and grey-level intensity values were used to describe image regions. In order to apply this technique to colour images, these images first had to be converted to greyscale. As a key objective of their work, Chen and Chen [26] strove to create a segmentation method which used both colour and texture information together. Chen and Chen extended the work of Ojala and Pietikainen, but rather than use contrast as one of the measured image region features, the authors used pixel colours. To do so, it was first necessary to reduce the number of colours in the image, using colour quantization. As a further enhancement to the Ojala-Pietikainen method, Chen and Chen replaced the use of local binary patterns with local edge patterns. Two region descriptors were then used simultaneously to describe each image region. The local edge patterns described the spatial structure of the local texture, and the colour feature described the colour distribution in each textured region. Also, in their paper Ojala and Pietikainen suggested that use of the G statistic to determine similarity of texture regions could be replaced by another method, such as the histogram intersection technique described by Swain and Ballard [28]. In their implementation, Chen and Chen opted to use this histogram intersection method.

For the colour quantization step, Chen and Chen chose to implement the adaptive HSV colour space quantization method proposed by Lo and Chen [29]. Chen and Chen chose to apply the Lo and Chen quantization technique to RGB colour space instead of the HSV colour space, although their paper provides no details on how RGB space was initially segmented, or the number of subregions created by each division of an RGB colour region.

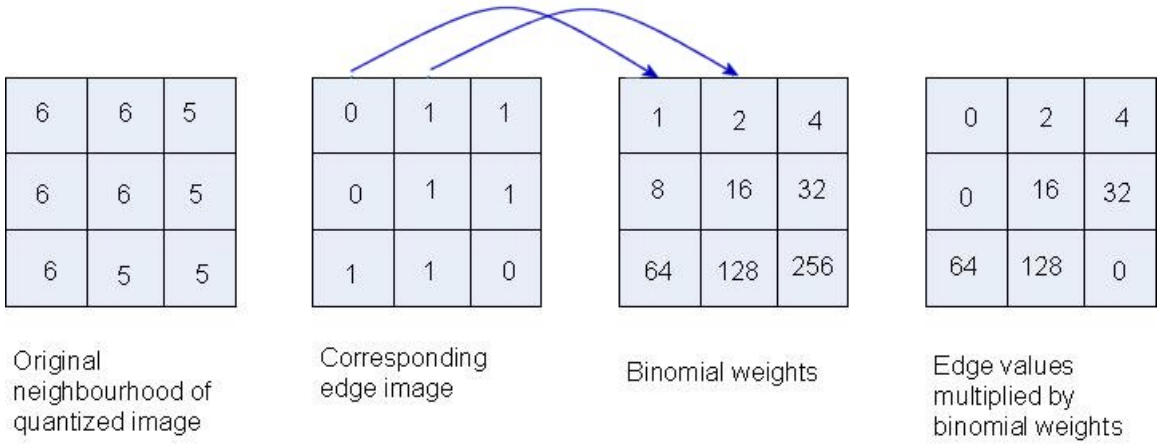
In the Chen-Chen method, once colour quantization of the image was complete, a colour histogram was constructed for each image region, using the quantized image colours in that

region. This colour histogram was then used as one of two texture descriptors for each region. The normalized colour histogram, *hc*, for each region was computed as follows:

$$hc_i = n_i / N \quad i = 0, \ldots, k\text{-}1 \tag{4.5}$$

where k is the total number of quantized colours, n_i is the number of pixels with colour label i in the region, and N is the total number of pixels in the region.

The second texture descriptor was based on local edge patterns, a modification to the local binary pattern descriptor implemented by Ojala and Pietikainen. This feature was calculated from an edge image derived from the colour quantized image. The binary edge values within each 3x3 neighbourhood were multiplied by binomial weights, and the resulting values summed to obtain the local edge pattern (LEP) for the neighbourhood. This value was assigned to the centre pixel of the neighbourhood, and the LEP was calculated in this manner for each pixel of the region.



LEP = 2 + 4 + 16 + 32 + 64 + 128 = 246

Figure 4.15: Example of LEP calculation for a 3x3 neighbourhood, taken from Chen and Chen paper [26].

The region's LEP values were then summarized in a normalized LEP histogram, *he*, where:

$$he_i = n_i / N \quad i = 0, \ldots, 511 \tag{4.6}$$

where n_i is the number of pixels with LEP value i in the region, and N is the total number of pixels in the region.

In the Chen-Chen method, the Swain and Ballard histogram intersection technique [28], applied to both the colour and LEP histograms, was used to assess the homogeneity of image regions. The homogeneity measure for the colour histograms, Hc, for any two regions R_i and R_i' was computed as:

$$Hc = \sum_{0}^{k-1} \min(hc_i, hc_i') \tag{4.7}$$

and the homogeneity measure for the LEP histograms, He, for the same regions was computed as:

$$He = \sum_{0}^{511} \min(he_i, he_i') \tag{4.8}$$

For both homogeneity measures, the closer the value was to 1, the more alike the two region histograms were. A single homogeneity value for two regions was computed by taking a weighted average of the Hc and He values, as shown:

$$H = 0.6Hc + 0.4He \tag{4.9}$$

where the weights 0.6 and 0.4 were determined experimentally.

Hierarchical splitting of the image proceeded exactly as described in the Ojala-Pietikainen technique, with the homogeneity test for an image region based on the homogeneity value, H, instead of the G statistic. After an initial partitioning of the image into subregions of a specified size, each subregion was temporarily divided into quarters and the six pairwise values of the homogeneity measure, H, were calculated. The subregion was considered inhomogeneous, and was divided, if the ratio of the maximum to the minimum of these six H values exceeded a given threshold:

$$HRatio = H_{\max} / H_{\min} > H\_Threshold \tag{4.10}$$

In this case, the authors determined experimentally that the threshold value, $H\_Threshold$, should be 1.1. In choosing such a low threshold value, the authors were opting to over-split rather than under-split the image at this stage, and then allow the agglomerative merging step to compensate for this over-segmentation.

Chen and Chen also modified the agglomerative merging step. In their implementation, they chose to calculate the merger importance, MI_{CC}, for any two adjacent regions as the homogeneity measure, H, of the regions divided by the square root of the number of pixels in the smaller of the two regions.

$$MI_{CC} = H \times \frac{1}{\sqrt{p}} \tag{4.11}$$

where p = number of pixels in the smaller of the two regions

By using the reciprocal of p, merging in the Chen-Chen method proceeded in order of decreasing MI_{CC} value, rather than increasing MI_{OP} value as was used in the Ojala-Pietikainen technique. The square root of p was used instead of p itself, since the range of H values is only 0 to 1, whereas the range of G values used in the Ojala-Pietikainen method is large. Use of the square root of p decreased the effect of this value on the MI_{CC} quantity. As regions were merged, both the colour and LEP histograms of the two regions were combined to form the histograms of the newly merged region.

In the Chen-Chen technique, two stopping criteria were implemented for the agglomerative merging step. Region merging proceeded until either of these criteria was true:

$$MIRatio1_{cc} = MI_{cur} / MI_{\min} < MI\_Threshold\_1 \tag{4.12}$$

$$MIRatio2_{CC} = MI_{cur} / MI_{\max} < MI\_Threshold\_2 \tag{4.13}$$

where MI_{cur} is the current best merger importance value, MI_{min} is the lowest merger importance value of any previous merger, and MI_{max} is the largest merger importance value of any previous merger. The authors stated that in the majority of cases it was the $MIRatio1_{CC}$ stopping rule which was triggered first. The $MIRatio2_{CC}$ rule was included in order to successfully handle some natural scenes, where the homogeneity measures within textured regions, and those between textured regions, were relatively similar. Through experimentation, the authors determined that the threshold value *MI_Threshold_1* should optimally be 0.68, and *MI_Threshold_2* should be 0.09.

Pixelwise classification was also used as the final step in the Chen-Chen implementation in order to improve detected region boundaries. In this case, the optimal disc radius was

determined experimentally to be 5 pixels, and the maximum number of passes of the pixelwise classification algorithm was limited to 30.

4.3.3.2 Implementation of the Ojala-Chen method

The Chen-Chen extension of the Ojala-Pietikainen technique, hereafter referred to as the Ojala-Chen method, seemed particularly relevant for the analysis of the lung tissue samples of this project. Segmentation of the tissue sample images is a difficult problem. Textures within these images are non-uniform, and the number of textures and texture regions is not known in advance. Colour is an important feature of these images, and use of colour and texture simultaneously to detect image regions seemed like a promising approach. For this thesis project, the Ojala-Chen technique was used to segment image areas, which were then evaluated to detect fibrin and necrosis within the images. However, in hindsight, this one technique could have been used to detect all regions of interest, including alveoli and areas of inflammatory cells, as well as fibrin and necrosis.

In this implementation, some modifications were made to the Chen-Chen adaptation of the Ojala-Pietikainen method. Colour quantization was applied to the RGB colour cube, as Chen and Chen suggested. Each dimension of the 255x255x255 RGB cube was initially divided into three, yielding 27 original colour regions, each with sides of length 85. Image pixels were assigned to the closest colour region based on Euclidean distance to each colour region's centre point. Any colour regions to which no pixels were assigned were eliminated. This elimination step is particularly important for the analysis of the tissue sample images; since the majority of colours in the tissue images are within the pink and purple ranges, a large number of the initial colour regions remain empty. Elimination of these empty colour regions ensures a finer segmentation of the relevant RGB colour regions before the maximum number of colours is reached. Splitting of colour regions proceeds based on mean square error of the regions, as described in the Chen and Chen paper. In this implementation, each split divides all dimensions of a colour region by two, yielding 8 subregions. A maximum split depth of four is used, as suggested by Chen and Chen, thus producing subsequent colour regions with sides of size 42, 21 and finally 10. The maximum number of colour regions was set to 200, although in practice the maximum splitting depth for any of the original colour regions was the stopping criteria which

typically triggered first, resulting in significantly fewer colour regions than this arbitrary maximum.

For the merging step of colour quantization, merging of adjacent colour regions proceeds until the minimum distance between the representative points of any two colour regions is at least 21. Initially, the representative point of each colour region is the centre point of that region. However, as regions of different sizes are merged, the concept of the centre point becomes obscure. Also, the number of pixels assigned to each of the regions involved in the merger should be taken into consideration when computing the exemplar for the newly merged region. Ideally, the average of all actual pixel values assigned to the regions involved in the merger would be calculated, and used as the new exemplar. However, as an efficiency improvement to this, when merging regions A and B, the new exemplar is calculated as:

$$[\ (\text{exemplar of region A}) \times (\text{number of pixels assigned to A}) + (\text{exemplar of region B}) \times (\text{number of pixels assigned to B}\ ]\ /\ (\text{total number of pixels in both regions}) \quad (4.14)$$

Hierarchical splitting was implemented as described in the Chen and Chen paper, with some modifications. In the implemented hierarchical splitting step, the 2000x2000 pixel images were initially split into subregions of size 100x100 pixels. The decision to perform this initial image partitioning before applying any homogeneity tests to the image subregions was implemented as an efficiency improvement to speed up overall processing time. It is extremely unlikely that this step would contribute to any over-segmentation of the image. The tissue sample images contain numerous small texture regions, which tend to be much smaller than 100x100 pixels, and assessing homogeneity of larger initial image regions is likely to result in large regions being incorrectly classified as homogenous due to a relatively even distribution of smaller texture regions within them. Rather than being a problem, the initial segmentation should improve the final splitting result, as well as speeding up the processing. Splitting was permitted to continue until a minimum block size of 12x12 pixels was reached, or until all regions were considered homogenous (Figure 4.16). Weightings applied to calculate the combined homogeneity measure of image regions were also altered from those used by Chen and Chen. Through experimentation with tissue images, it was discovered that placing less emphasis on colour distributions, and more on edge patterns, produced better results. An

optimal weighting for these images was determined to be:

$$H = 0.4Hc + 0.6He \tag{4.15}$$

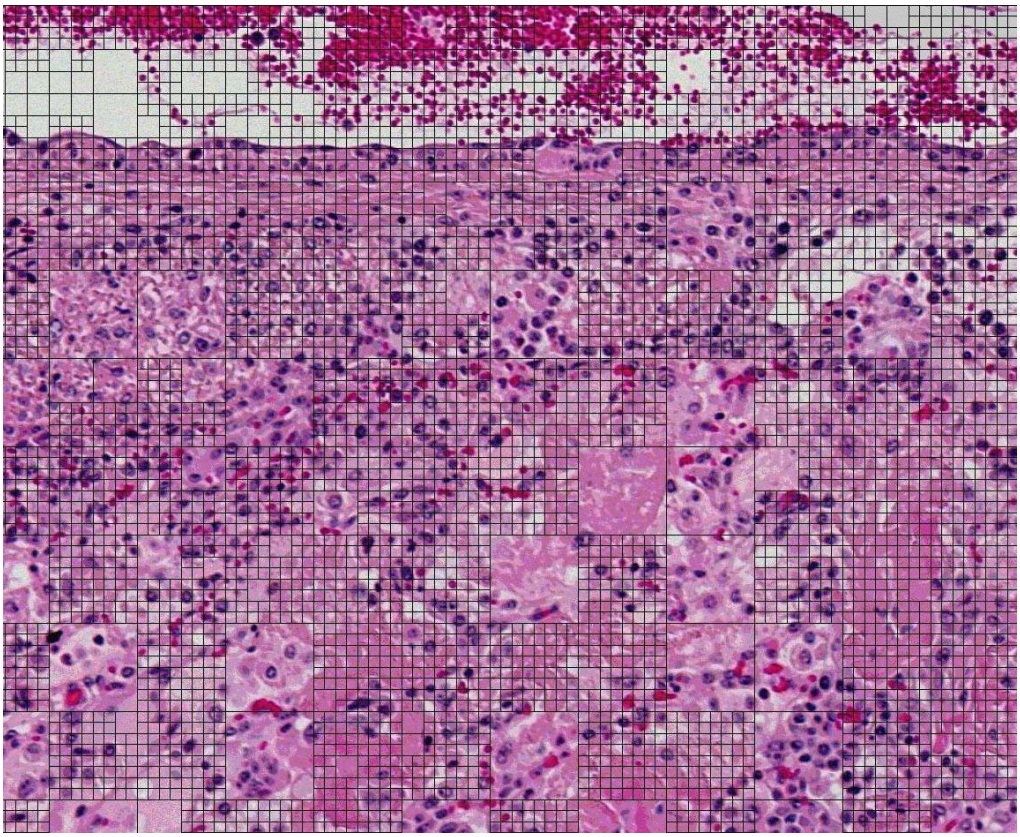


Figure 4.16: Hierarchical splitting of a sample image

For the agglomerative merging step, the threshold for the stopping criterion $MIRatio2_{CC}$ was changed to *MI_Threshold_2* = 0.10, from the 0.09 value proposed by Chen and Chen. The optimal value for stopping criterion $MIRatio1_{CC}$ was determined to be *MI_Threshold_1* = 0.8. Both values were determined experimentally.

Completion of the colour quantization, hierarchical splitting and agglomerative merging steps produces a segmented image in which regions of essentially uniform colour and texture have been highlighted (Figure 4.17). The next step in both the original Ojala-Pietikainen technique, and in the Chen-Chen extension to it, is pixelwise classification in order to improve the accuracy of image region boundaries. This final step was not included in the implementation of the Ojala-Chen technique for this project due to time limitations. As a result, image regions produced by the implemented Ojala-Chen method have jagged, square boundaries.

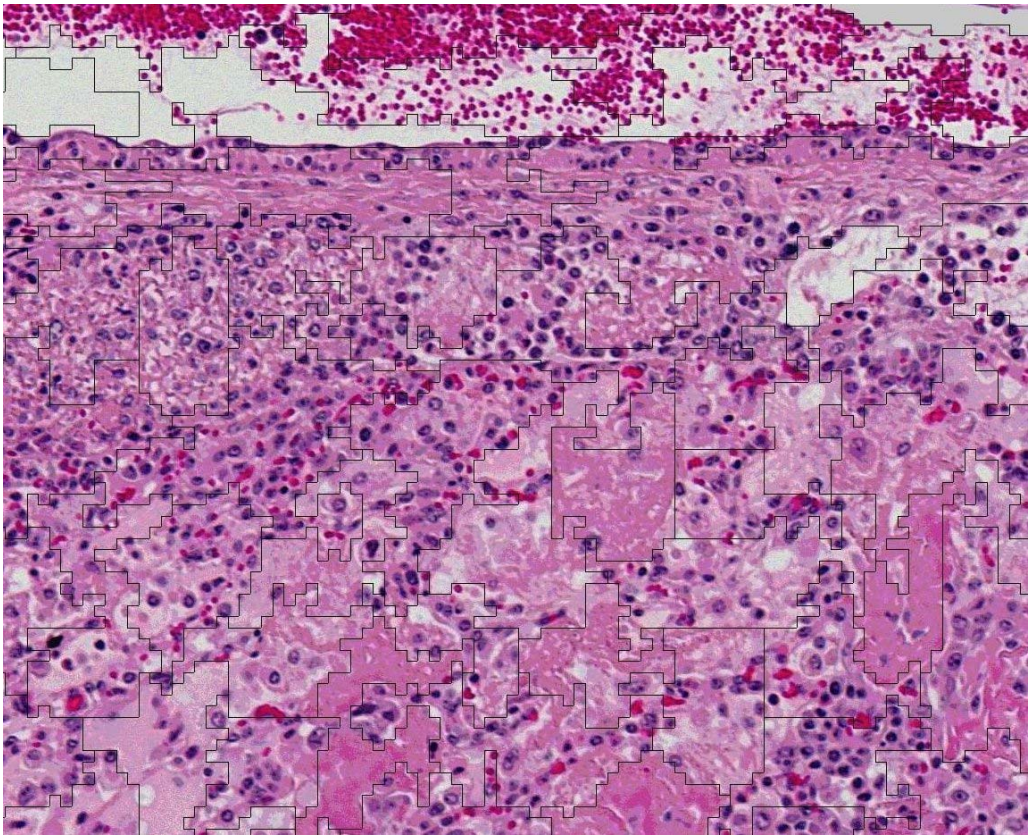


Figure 4.17: Results of agglomerative merging

Implementation Alternatives

Matlab offers two colour quantization algorithms: uniform quantization and minimum variance quantization. Uniform quantization divides the colour cube into equal blocks at each step. This is equivalent to the first step in the implemented colour quantization technique. Matlab's minimum variance quantization divides the colour cube into regions which are not uniform in size or shape, but which are based on the actual distribution of image pixels to the regions. Areas of the colour cube where no pixels exist are eliminated. This type of quantization is the aim of the colour quantization technique described by Chen and Chen, and implemented in this project. Testing using Matlab's minimum variance quantization did show a dramatic increase in processing speed over the implementation of the Chen-Chen colour quantization method. The disadvantage of Matlab's minimum variance quantization algorithm, however, is that the user must specify at the outset the number of colour regions to be created. With the colour quantization technique implemented for this project, the final number of colour

regions is not pre-set but is determined by the distribution of pixel colours, and will vary from image to image. Another advantage of the implemented colour quantization technique is that background (non-tissue) regions of each image are ignored. These background areas are not quantized and thus do not occupy colours in the limited quantized colour list. Consequently, it was decided that the colour quantization method described by Chen and Chen would be used, instead of utilizing one of Matlab's built-in quantization functions.

Matlab's *quadtree* decomposition algorithm was examined as a potential means of implementing the hierarchical splitting step. However, Matlab's *quadtree* decomposition function works only on greyscale images, and colour information is an important feature when deciding whether regions should be split in the Ojala-Chen technique. As a result, Matlab's *quadtree* decomposition was not used.

4.3.3.3 Determining image regions which correspond to fibrin and necrosis

The image segmentation produced by the implemented Ojala-Chen method could be used to partition regions corresponding to any image features of interest. For the thesis project, this technique was applied only to detect areas of fibrin and necrosis. Once the image segmentation via the Ojala-Chen method was complete, statistics calculated from co-occurrence matrices were used to classify image regions as fibrin, necrosis, or "other".

The number of image regions to be evaluated was reduced by initially selecting only those regions which have the majority of their pixels within the dark pink colour range, the colour region which is characteristic of both fibrin and necrosis. To determine the colour categorization of pixels in each region, the overlap of the segmented image with the RGB clustered image produced in the detection of alveoli step (refer to section 4.3.1 for details) was found. The pixels within the RGB clustered image which corresponded to the pixels within each of the image regions were located, and their colours noted. Image regions which had a minimum of 57% of their pixels categorized as "pink" in the RGB clustered image were considered potential fibrin or necrosis areas, and were selected for further evaluation. The "pink" category in the RGB clustering images was intended to group areas corresponding to fibrin, necrosis and alveolar walls. The 57% threshold was chosen experimentally through sampling of various image regions.

Matlab's *graycomatrix* function was used to compute the co-occurrence matrices (as described in section 3.2.7). Input to *graycomatrix* is a rectangular image, requiring that each image region of interest be isolated as a separate subimage. The bounding box of each image region of interest was determined in order to find these subimages. Since segmented image regions produced by the Ojala-Chen method are rarely rectangular, this required step introduced additional non-region pixels into the subimage submitted to the *graycomatrix* function. In order to exclude these pixels from the co-occurrence matrices, the following steps were taken. Background (non-image region) pixels within these subimages were labeled as black, with an RGB value (0,0,0). This is unlikely to cause any overlap with true image region pixels, since "black" pixels within the image rarely, if ever, have a precise (0,0,0) RGB value. Calculation of the co-occurrence matrices was allowed to proceed with this altered rectangular image region. The first row and first column of the resulting co-occurrence matrices were then eliminated, since this row and column contain the co-occurrence values for the lowest pixel value (0,0,0) with all other pixel values in the quantized input image.

The *graycomatrix* function requires that the image regions provided as input are greyscale images. The image regions were converted to greyscale using Matlab's *rgb2gray* function. This colourspace conversion yields greyscale images consisting of 256 intensity values. The *graycomatrix* function, by default, further reduces this to eight intensity values only, resulting in co-occurrence matrices of size 8x8.

Co-occurrence matrices for the four angular directions 0, 45, 90 and 135 degrees were created, for distances of both 1 and 3. The texture features contrast, correlation, energy, and homogeneity were then calculated for each direction and each distance. Feature values were averaged across the four directions for each distance, producing a total of eight values which were then used for classification of the regions.

The required output of the classification is three categories of regions from those processed in this analysis. These categories are fibrin, necrosis, and "other", a catch-all category to group alveolar wall areas and any other type of tissue component. Each of the four texture features were first examined individually in order to determine whether any of them produced value ranges which were unique for each of the three categories of region. With contrast, using either distance, the value ranges for fibrin, necrosis and other areas all overlapped. A similar result was obtained for homogeneity. With correlation, at both distances, necrosis regions

produced a unique range of values, but the ranges for fibrin and other areas showed some amount of overlap. With energy calculated for distance 1, the range of values of fibrin overlapped with those of other areas, as did the range of values for necrosis. The necrosis and fibrin ranges were distinct from one another, however, provided that a single anomalous result was omitted. For energy calculated for distance 3, the ranges of values for fibrin, necrosis and other areas all overlapped. Consequently, correlation could be used as a single feature to distinguish necrosis from any other area. There was, however, no single feature which accurately distinguished fibrin areas from other types of area.

Classification of image regions using all four feature values, calculated at distance 1, was also attempted. In this case, Matlab's *classify* function was used, which is an implementation of discriminant analysis. The training set consisted of 32 image regions: 11 fibrin areas, 6 necrosis areas and 15 "other". Classification results were tested using the "leave one out" methodology. Results using classification based on all four texture features proved to be less accurate than those obtained using correlation alone. For the final implementation, correlation calculated at distance 1 was used to categorize the image regions as fibrin, necrosis or "other". A sample image with regions of fibrin and necrosis highlighted is shown in Figure 4.18.

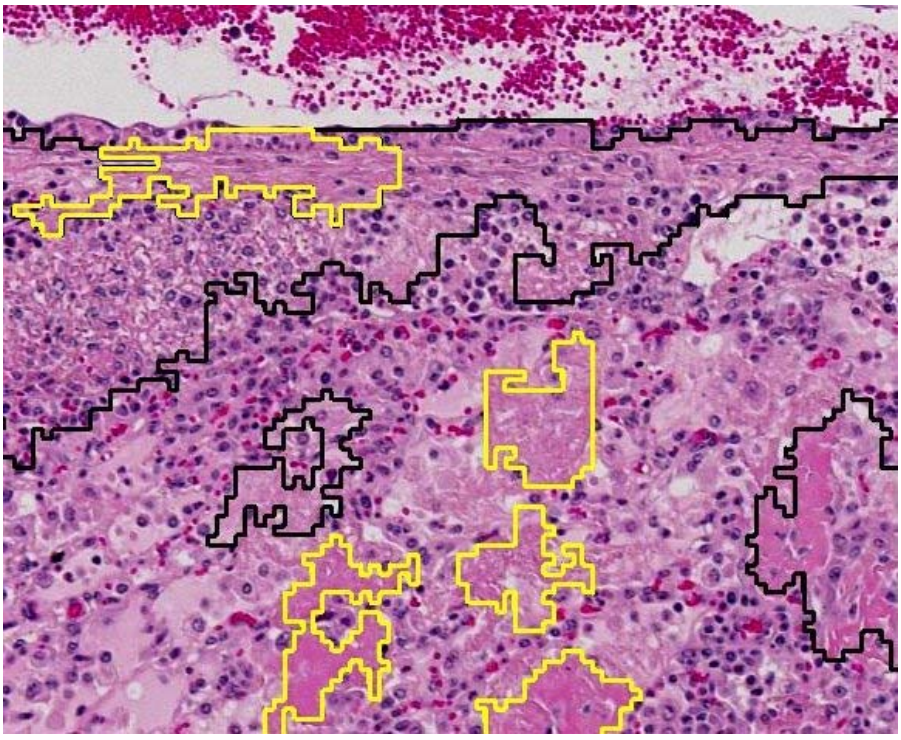


Figure 4.18: Regions identified as fibrin (outlined in black) and necrosis (outlined in yellow)

CHAPTER 5
RESULTS

5.1 Assessment of results

The thesis software produces a total of four highlighted images for each slide region analyzed. Alveolar boundaries are marked on two highlighted images: one which shows all alveoli whether fluid-filled or air-filled, and a second which highlights only those alveoli which are considered fluid-filled. The inflammatory cells analysis produces one image on which all areas of inflammatory cells are outlined. The final image outlines boundaries of areas of fibrin and necrosis, using two different colours to distinguish these areas from one another. Two reports are also produced. The "Image Analysis Report" summarizes information about alveoli and inflammatory cells in the image, including the percentage of alveoli which are fluid-filled, the percentage of total alveolar space which contains fluid, and the percentage of tissue area which contains regions of inflammatory cells. The "Fibrin and Necrosis Report" lists the percentage of tissue area which contains fibrin and that which contains necrosis.

Results of the programs were assessed by comparing the disease-indicating areas found by the automated analysis with corresponding areas that were assessed manually by a pathology expert. Dr. G. Wobeser identified all alveolar boundaries, areas of inflammatory cells, fibrin and necrosis by manually outlining these areas for each tissue sample image. For each image, Dr. Wobeser marked the boundaries and labeled the regions containing these disease-indicating features on a transparent film laid on top of a printed copy of the image. To compare results of the expert manual analysis and the automated analysis, the highlighted images produced by the programs were also printed. The marked transparency was placed over each of the four highlighted images for each tissue sample, the regions identified by both analyses were manually compared and the results tabulated.

Analysis results were extremely varied. In some cases, particularly with alveoli detection, region boundaries identified by both analyses were almost identical. In other cases, a single region identified by Dr. Wobeser as containing a disease-indicating feature overlapped with several smaller areas of the same type identified by the automated analysis. Conversely, in

some instances a larger area identified by the automated analysis overlapped with one or several smaller areas of the same type identified by Dr. Wobeser. There were also numerous instances in which the automated analysis did not identify any disease indicators in areas corresponding to one identified by Dr. Wobeser, and cases where the automated analysis identified disease-indicating features where none were marked in the manual analysis. In all cases, the expert manual analysis results produced by Dr. Wobeser were considered the true standard by which the automated results were evaluated.

Sensitivity and specificity are typically the measurements by which such diagnostic results are assessed. To determine these values, results are counted as true and false positives, and true and false negatives, where:

- A true positive (TP) indicates an actual case identified as positive by the analysis
- A false negative (FN) indicates an actual case which was not identified by the analysis
- A true negative (TN) indicates a nonexistent case that was correctly identified as negative
- A false positive (FP) indicates a nonexistent case that was incorrectly identified as positive

		Actual Case	
		Positive	**Negative**
Analysis Result	**Positive**	TP	FP
	Negative	FN	TN

Figure 5.1: Scoring results

Sensitivity measures the percentage of actual cases found.

$$\text{Sensitivity} = TP / (TP + FN) \tag{5.1}$$

Specificity measures the percentage of actual non-cases found.

$$\text{Specificity} = TN / (TN + FP) \tag{5.2}$$

A receiver operating characteristic (ROC) curve is created by plotting sensitivity vs. specificity for a set of tests. An ROC curve illustrates the tradeoff between increasing sensitivity at the expense of specificity.

Standard ROC curves cannot be used in an analysis such as this, however, since it is not possible to count true negatives. Counting true negatives would require that there is a definite set of objects containing a countable number of members. Each member of the set is either positive for a disease-indicating feature, or negative for it. However, with the images, the number of regions that can possibly be identified is unlimited. Boundaries and sizes of regions can be marked in any number of ways, creating an unpredictable number of objects. If members of the set were limited to only those identified by the manual analysis, it would not be possible to accurately count false positives, since these regions do not necessarily correspond to any identified region. There are an unlimited number of possible negative areas that could be incorrectly identified as positive. Consequently, determining specificity for this type of analysis is not possible. (There is one exception, as discussed in section 5.3). Instead of plotting a traditional ROC curve, a free-response ROC curve is used, which plots sensitivity vs. the number of false positives. In this case, the goal is to maximize sensitivity while minimizing the number of false positives. Free-response ROC curves are typically used for assessing medical image analysis results [30; 31].

True positive, false negative and false positive scores are counted for each disease indicator for each of the analyzed slide areas, and sensitivity values are computed. Since it is not possible to calculate specificity, Positive Predictive Value is determined instead. Positive Predictive Value measures the percentage of detected areas that are correct.

$$\text{Positive Predictive Value} = TP / (TP + FP) \tag{5.3}$$

Scoring the automated results presented a number of additional challenges, such as instances where there were significant differences in the boundaries of identified regions but also significant region overlap. It was necessary to define rules for scoring the results. The following standards were used:

- If an area identified by Dr. Wobeser overlapped with any area or areas marked by the automated analysis, this region was considered detected and counted as a True Positive. For this count, accuracy of the region boundaries was not considered.

- The number of regions to be counted was the number of areas identified by Dr. Wobeser. Consequently, if several smaller areas of necrosis in the automated analysis result corresponded to a single area of necrosis in the manual analysis, for example, this was counted as one True Positive.

- Each incorrectly identified region was counted as one False Positive. If three areas of fibrin were wrongly marked within a single region identified as necrosis, for example, this was scored as three false positives for fibrin. (As previously discussed, the number of false positives is unlimited).

- In some cases, the entire tissue area was identified by Dr. Wobeser as containing a certain disease indicator, without any specific regions of this feature marked. This posed a problem for scoring analysis results which did outline specific areas. In these cases, the number of areas of this feature identified by the automated analysis is not meaningful, and so no count was done. Instead, an attempt was made to assess the analysis results for these images based on total area occupied by the feature in the manual and automated analysis results.

- Areas in some of the slides were marked by Dr. Wobeser as "not to be analyzed". Regions consisting primarily of tissue other than lung parenchyma are not typically examined by pathologists in order to diagnose the disease state present, and so were labeled in this way (see section 4.1 for details). However, these regions were not excluded from the automated analysis. No intelligence was built into the thesis software to identify and skip areas such as these; consequently, the algorithms did attempt to identify disease features in these regions. Any disease features identified in "not to be analyzed" regions were counted as false positives. It could be argued that results in these regions should not be counted, however. An "Adjusted False Positive" value is thus also included for each count. This adjusted figure excludes any false positives which were located in "not to be analyzed" areas.

Results for each of the three stages of the automated analysis (detection of alveoli, detection of areas of inflammatory cells, detection of fibrin and necrosis) for each of the 50 analyzed slide regions are summarized in separate sections, which also include an examination of these results. For a few sample images, the highlighted images created by the automated analysis and the detailed results tables for these images are included in section 5.2. This amount of detail is not included within the thesis body for all 50 images. However, the full set of highlighted images and detailed results are included in the supplementary information accompanying the thesis (see Appendix A).

5.2 Sample Detailed Analysis Results

5.2.1 Image Region [12001, 20001] of Slide D03-10538

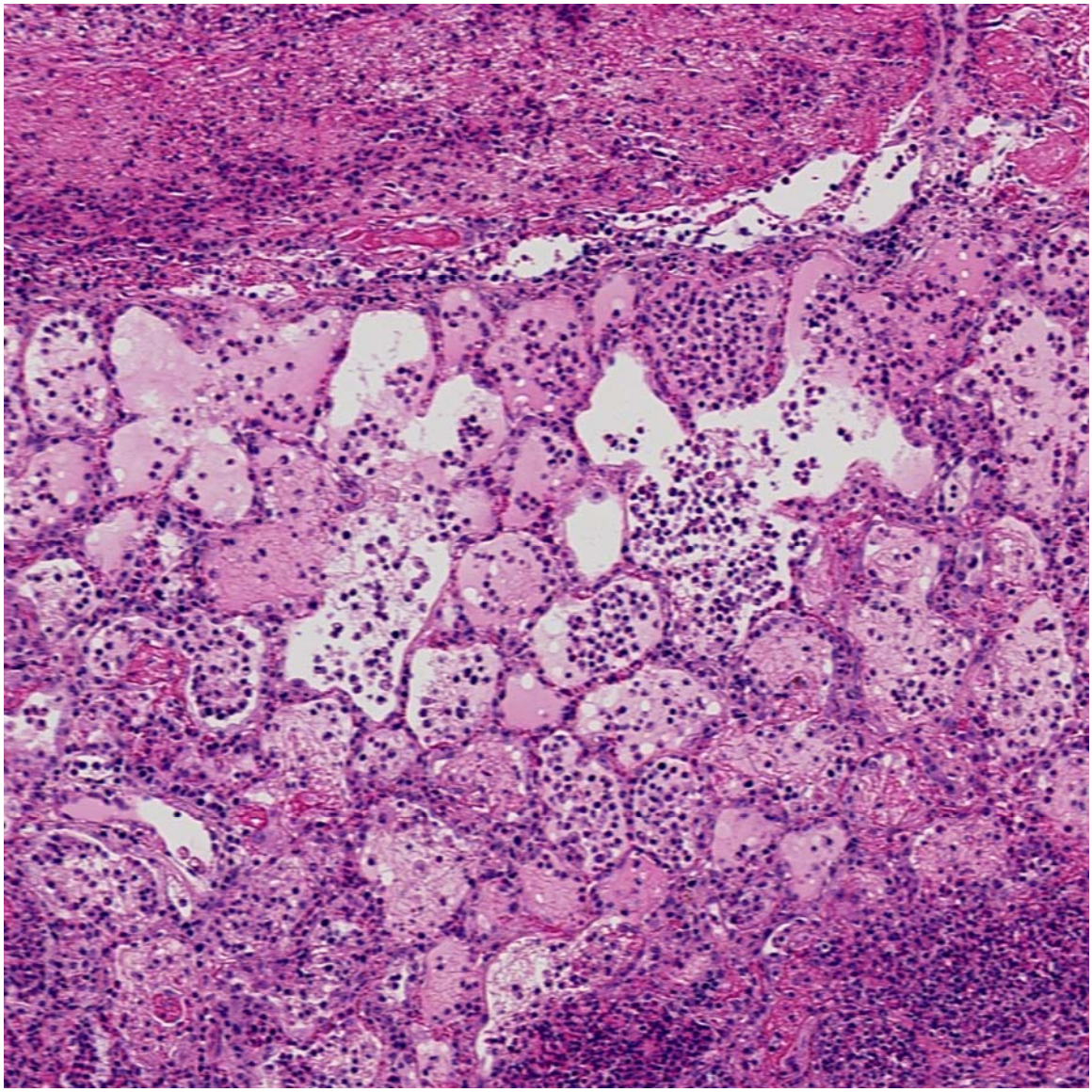


Figure 5.2: Region [12001,20001] of slide D03-10538

5.2.1.1 Automated Analysis Results

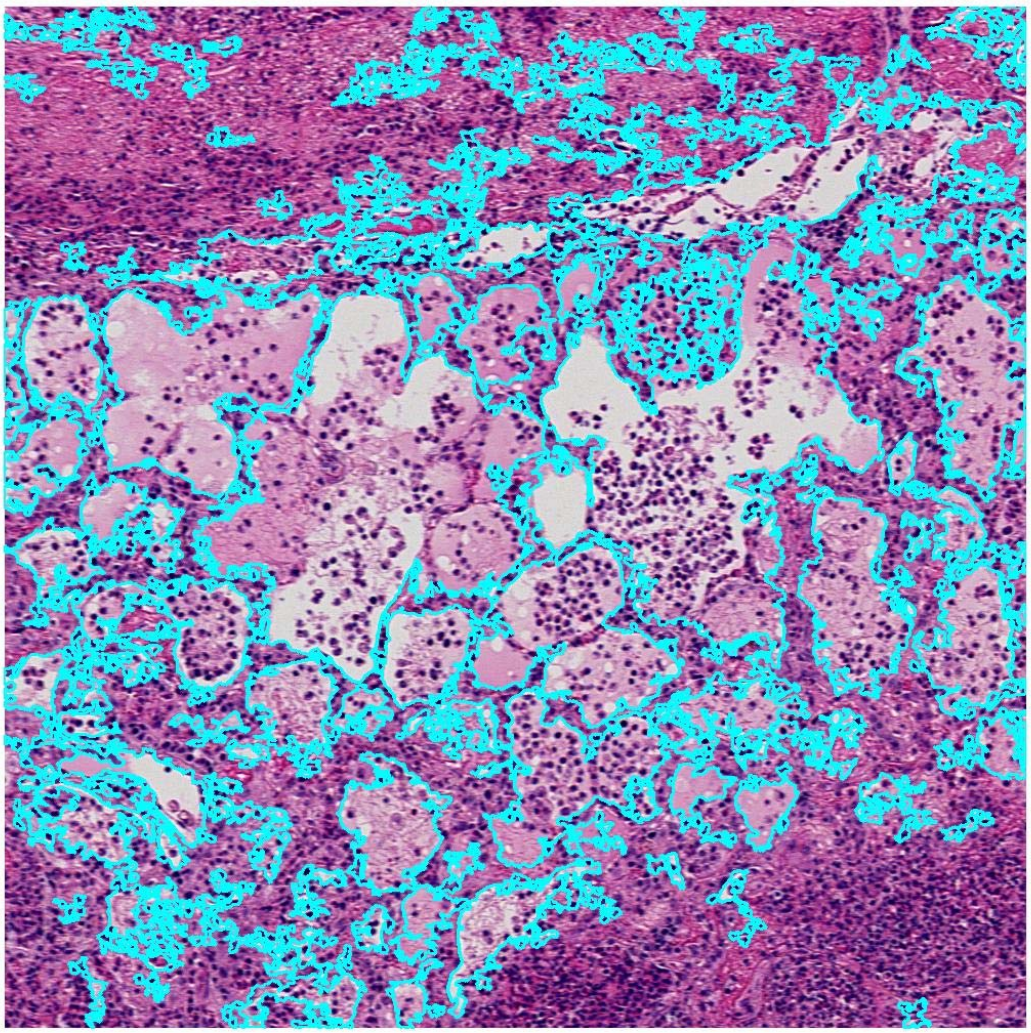


Figure 5.3: All alveoli (air-filled and fluid-filled)

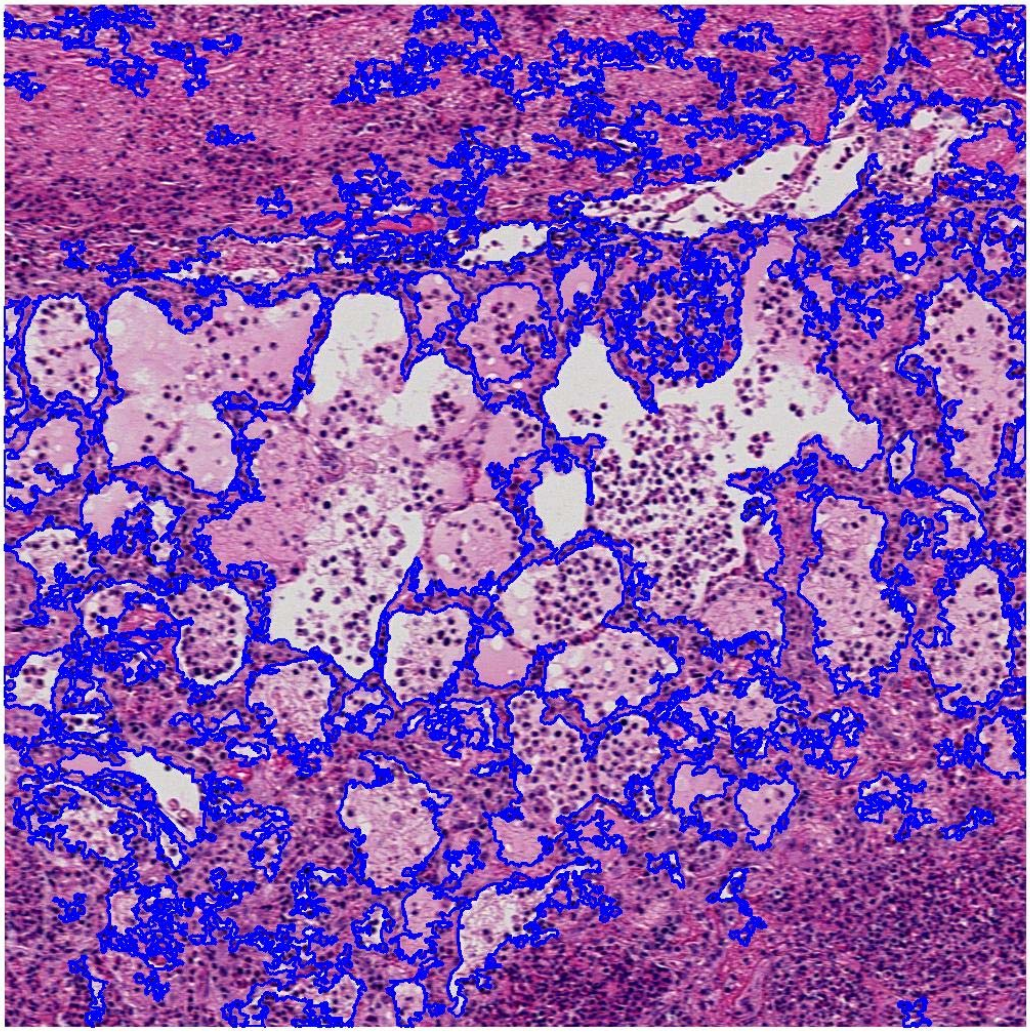


Figure 5.4: Fluid-filled alveoli

```
**********  Image Analysis Report **************************
Image : D0310538_12001x20001
Processing start time : 11-May-2006 14:08:11

Alveoli and Alveolar Contents:
Number of alveoli in image : 111
Percentage of alveoli which are considered fluid-filled (contain more than 33 percent fluid) : 100.00
Total alveolar space in image (pixels) : 1881585
Percentage of total alveolar space which is fluid-filled :  68.75


Areas of Inflammation:
Percentage of tissue area which is considered inflammation :  36.21
```

Figure 5.5: Image Analysis Report

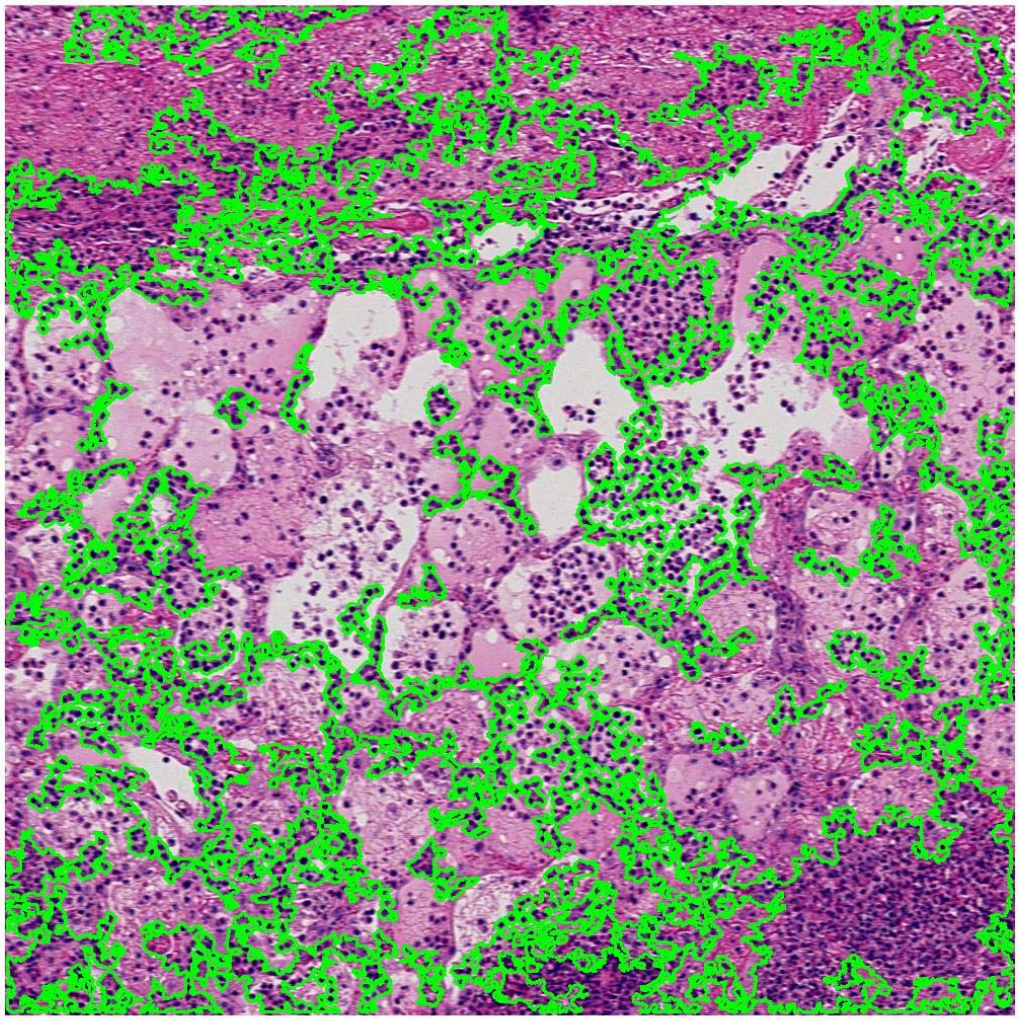


Figure 5.6: Areas of inflammatory cells

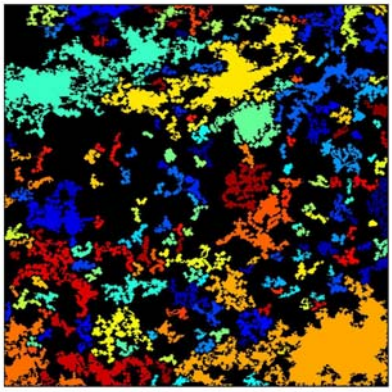


Figure 5.7: Areas of inflammatory cells shown as coloured regions

(Figure 5.7 is provided as an aid in interpreting the regions outlined in the Figure 5.6. Colours assigned to regions in this image are not significant.)

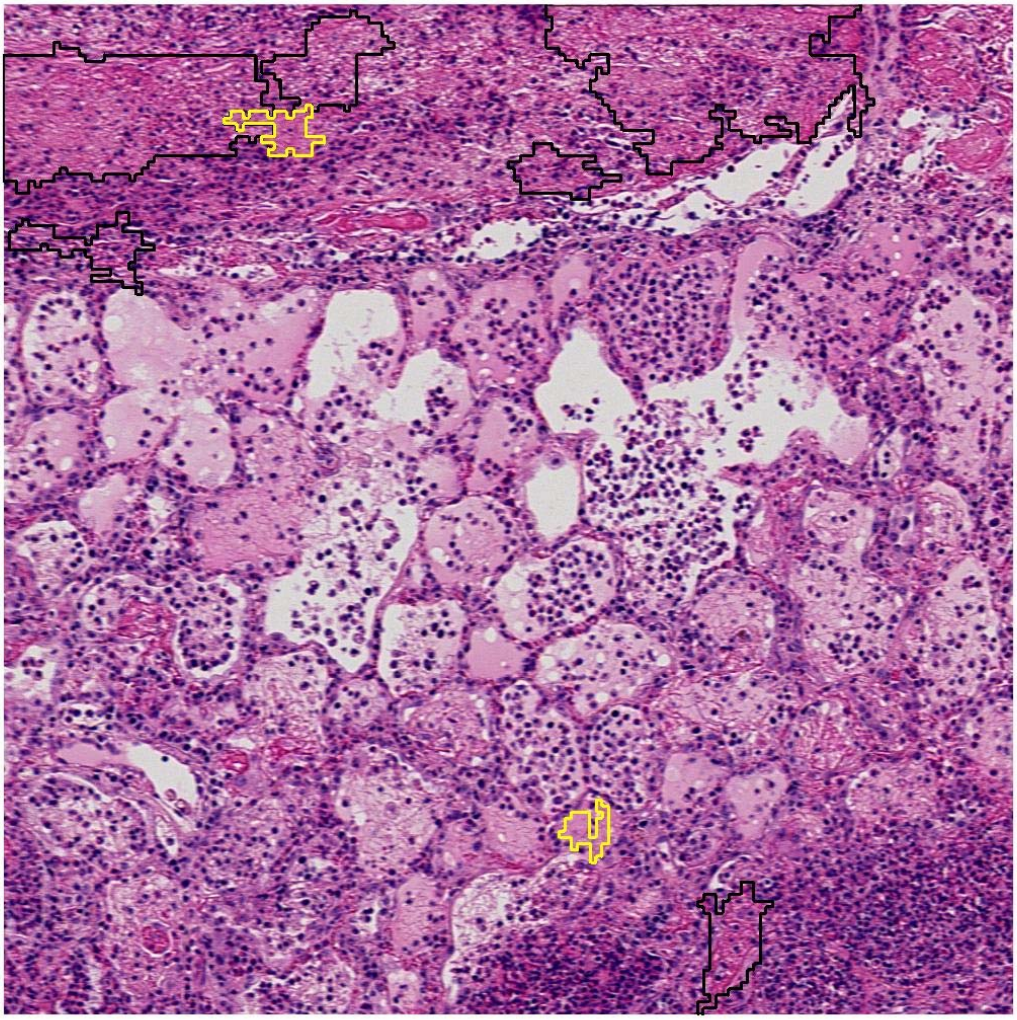


Figure 5.8: Areas of fibrin (outlined in black) and areas of necrosis (outlined in yellow)

********** Fibrin and Necrosis Report **********
Image : D0310538_12001x20001
Processing start time : 11-May-2006 14:46:55

Percentage of tissue area which is considered fibrin : 8.28

Percentage of tissue area which is considered necrosis : 0.37

****

Figure 5.9: Fibrin and Necrosis Report

5.2.1.2 Manual Analysis Results

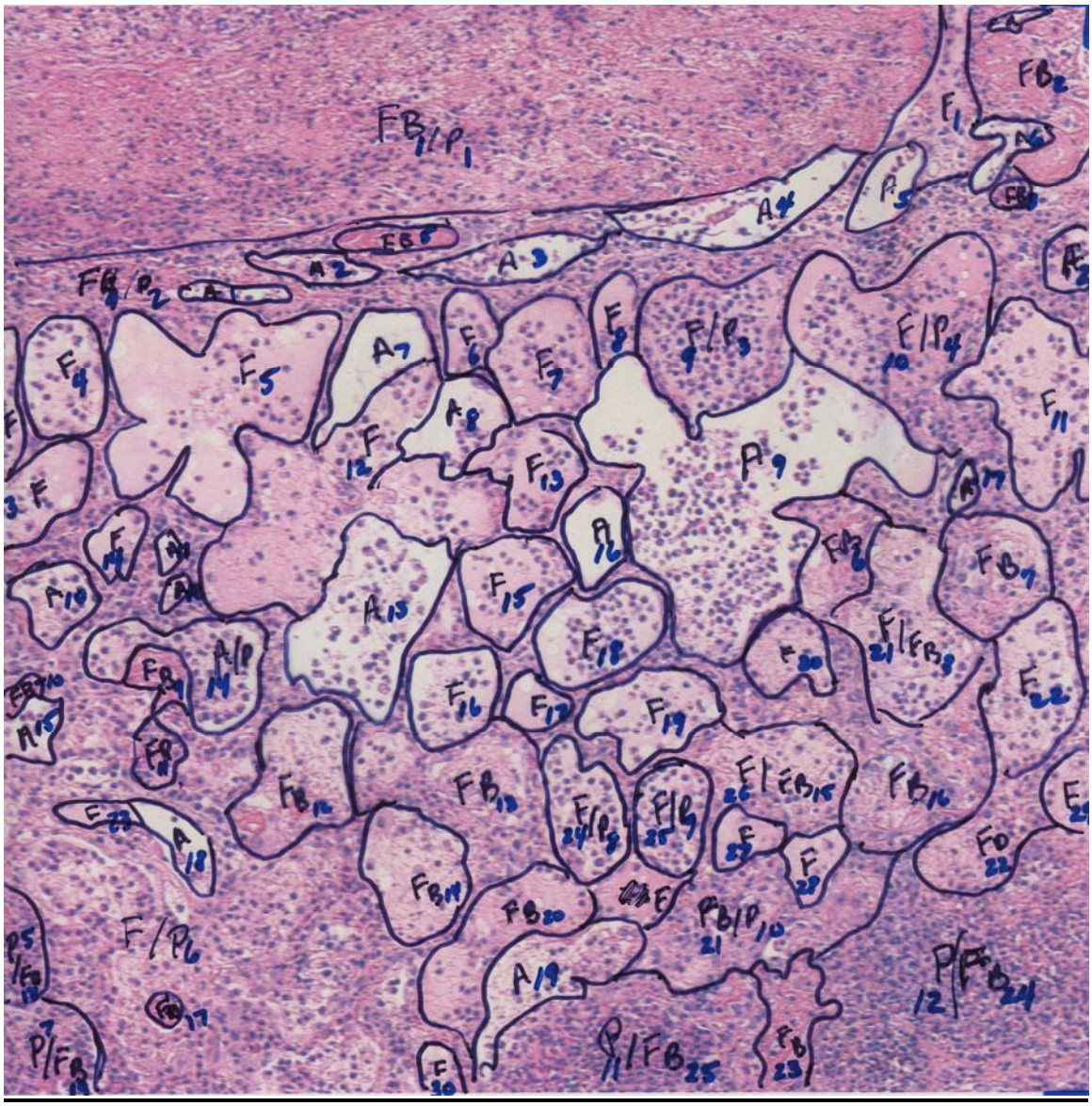


Figure 5.10: Expert manual analysis of region [12001, 20001] of slide D03-10538

Legend:	
A	Air-filled alveolus
F	Fluid-filled alveolus
P	Area of inflammation
FB	Fibrin
N	Necrosis

5.2.1.3 Comparison of Results

The image below (Figure 5.11) shows all air-filled and fluid-filled alveoli identified by the manual analysis overlaid on the fluid-filled alveoli image produced by the software. Since the software identified all alveoli in this image as fluid-filled, this one image shows the complete set of alveoli detected by the programs. A detailed comparison of the manual and automated results for alveoli detection is included in Table 5.1.

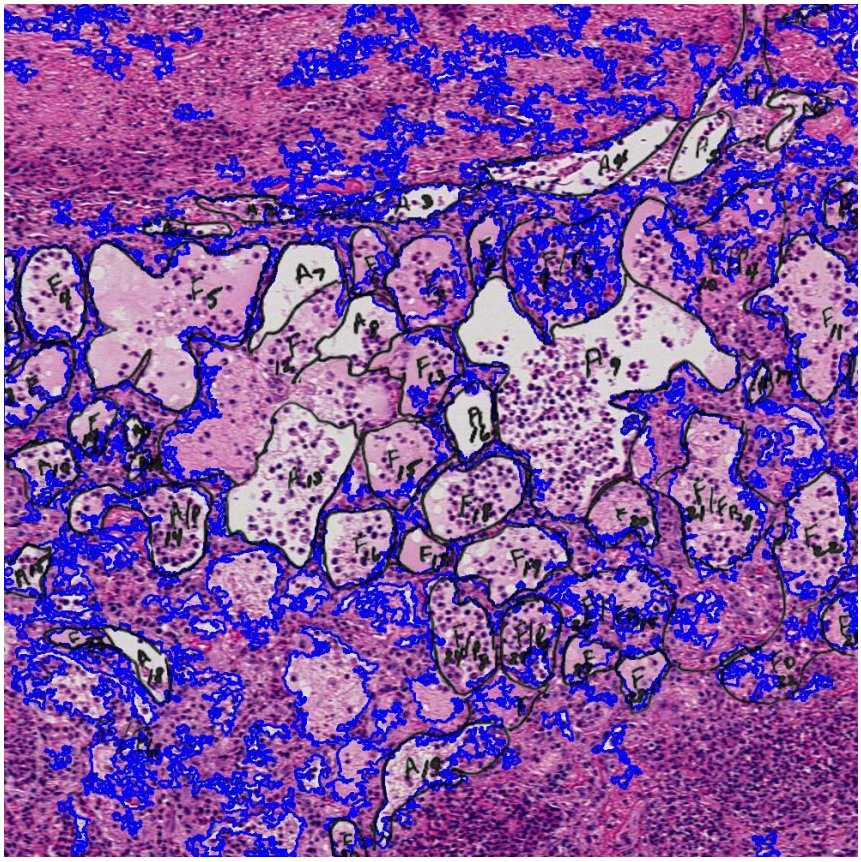


Figure 5.11: Composite image comparing alveoli detected by the manual and automated analyses

Table 5.1: Detection of all alveoli

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
A1	Identified as part of larger alveolus, along with A2.	TP
A2	Identified as part of larger alveolus, along with A1.	TP
A3	Found.	TP
A4	Identified as part of larger alveolus, along with A5, A6.	TP

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
A5	Identified as part of larger alveolus, along with A4, A6.	TP
A6	Identified as part of larger alveolus, along with A4, A5.	TP
A7	Identified as part of larger alveolus, along with A8, F12, F13 A13, F15	TP
A8	Identified as part of larger alveolus, along with A7, F12, F13, A13, F15.	TP
A9	Identified as part of larger alveolus, along with F10, A16, F20.	TP
A10	Identified as part of larger alveolus, consisting of A10, F14, A11, A12.	TP
A11	Identified as part of larger alveolus, consisting of A10, F14, A11, A12.	TP
A12	Identified as part of larger alveolus, consisting of A10, F14, A11, A12.	TP
A13	Identified as part of larger alveolus, along with A7, A8, F12, F13, F15.	TP
A14	Found. Excellent match on boundaries.	TP
A15	Found. Excellent match on boundaries.	TP
A16	Identified as part of larger alveolus, along with F10, A9, F20.	TP
A17	Found. Excellent match on boundaries	TP
A18	Identified as part of larger alveolus, along with F23.	TP
A19	Found. Excellent match on boundaries.	TP
	All air-filled alveoli (A1 – A19) identified as fluid-filled	FP x 19
F1	Found. Shown as part of larger alveolus	TP
F2	Found. Good match on boundaries	TP
F3	Found. Shown as two separate fluid-filled alveoli	TP
F4	Found. Excellent match on boundaries	TP
F5	Found. Excellent match on boundaries	TP
F6	Found. Good match on boundaries	TP
F7	Found. Excellent match on boundaries	TP
F8	Found. Good match on boundaries	TP
F9	Found. Jagged boundaries	TP
F10	Parts found. Overlaps two smaller areas	TP
F11	Found. Excellent match on boundaries	TP
F12	Found. Considered part of one large alveolus, along with: A7, A8, F13, A13, F15	TP
F13	Found. Considered part of one large alveolus, along with: A7, A8, F12, A13, F15	TP
F14	Found. Considered part of one larger alveolus along with A11, A12	TP
F15	Found. Considered part of one large alveolus, along with: A7, A8, F12, A13, F13	TP
F16	Found. Excellent match on boundaries	TP
F17	Found. Considered part of larger alveolus, along with F18, F19	TP
F18	Found. Considered part of larger alveolus, along with F17, F19	TP
F19	Found. Considered part of larger alveolus, along with F17, F18	TP
F20	Found. Considered part of larger alveolus, along with A9	TP

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
F21	Found. Excellent match on boundaries	TP
F22	Found. Excellent match on boundaries	TP
F23	Found. Considered part of larger alveolus, along with A18	TP
F24	Found. Considered part of larger alveolus, along with F25	TP
F25	Found. Considered part of larger alveolus, along with F24	TP
F26	Found. Good match on boundaries	TP
F27	Found. Excellent match on boundaries	TP
F28	Found. Excellent match on boundaries	TP
F29	Found. Excellent match on boundaries	TP
F30	Found. Excellent match on boundaries	TP
	Small fluid-filled areas identified within areas of fibrin FB1, FB2, FB4, FB6, FB7, FB9, FB13, FB16, FB19, FB20, FB21, FB22, FB23, FB24.	FP x 14
	Fibrin areas FB5, FB12, FB14 identified as fluid-filled alveoli	FP x 3

<u>All Alveoli</u>
Number of alveoli (air-filled and fluid-filled) = 49
TP = 49 FN = 0 Sensitivity = 100.0%
FP = 17 Positive Predictive Value = 74.2%

<u>Fluid-Filled Alveoli</u>
Number of fluid-filled alveoli = 30
TP = 30 FN = 0 Sensitivity = 100.0%
Within set of all objects: FP = 36 Positive Predictive Value = 45.5%
Within set of all alveoli: FP = 19 Positive Predictive Value = 61.2%
TN = 0 Specificity = 0.0%

The image below (Figure 5.12) shows regions of inflammatory cells identified in the manual analysis overlaid on the highlighted inflammatory cells image produced by the software. Refer to table 5.2 for a detailed comparison of these results.

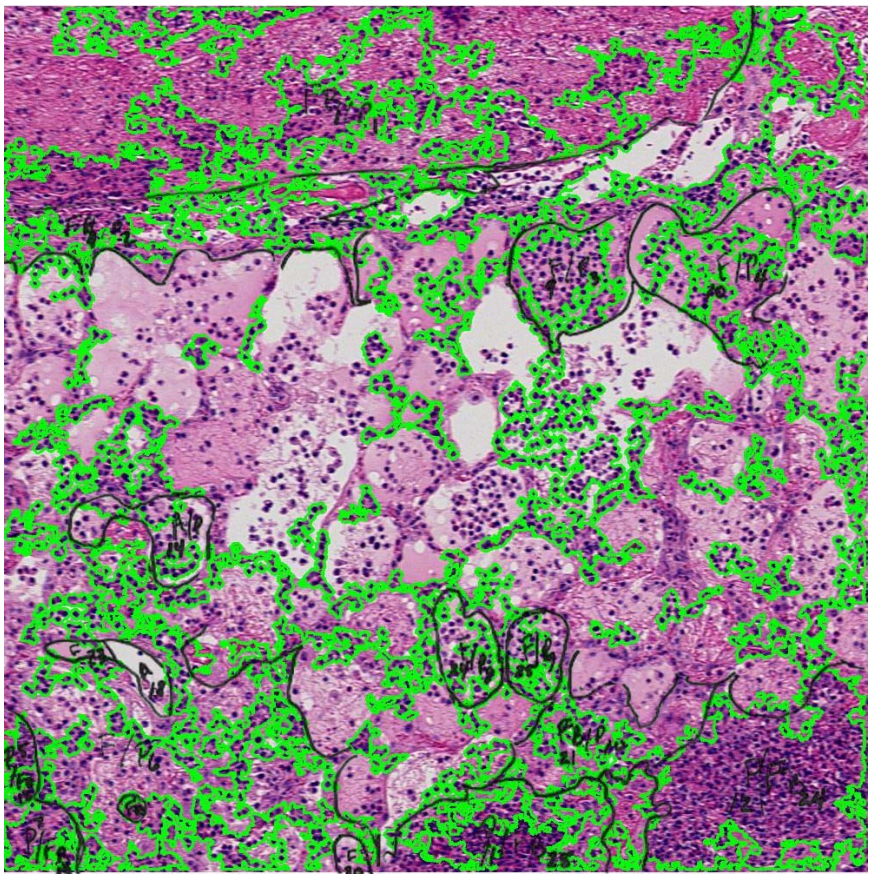


Figure 5.12: Composite image comparing regions of inflammatory cells identified by the manual and automated analyses

Table 5.2: Detection of areas of inflammatory cells

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
P1	Overlaps with numerous smaller identified areas	TP
P2	Found. Almost all of area identified in two separate regions.	TP
P3	Found. Good match on boundaries	TP
P4	Overlaps with five smaller identified areas.	TP
P5	Found. Identified as part of larger area, along with P7.	TP
P6	Overlaps with numerous smaller identified areas.	TP
P7	Found. Identified as part of larger area, along with P5.	TP
P8	Found. Identified as part of larger area, along with P8.	TP
P9	Found. Identified as part of larger area, along with P9.	TP

Region identified by Dr. Wobeser	*Corresponding region identified by programs*	*Count*
P10	Overlaps with three smaller identified areas.	TP
P11	Found. Good match on boundaries.	TP
P12	Found. Good match on boundaries.	TP
	Inflammation identified with alveoli A2, A4, A8, A9, A10, A13, F2, F4, F7, F8, F13, F19, F21, F22, F26	FP x 15
	Inflammation identified within fibrin areas FB2, FB7, FB10, FB11, FB12, FB13, FB16, FB20, FB22, FB23	FP x 10

Number of areas of inflammatory cells = 12
TP = 12 FN = 0 Sensitivity = 100.0%
FP = 25 Positive Predictive Value = 32.4%

The image below (Figure 5.13) shows regions of fibrin identified by the manual analysis overlaid on the fibrin and necrosis image produced by the software. For this particular image, no areas of necrosis were identified in the expert manual analysis. A detailed comparison of the manual and automated results for fibrin detection is included in Table 5.3, and for necrosis detection in Table 5.4.

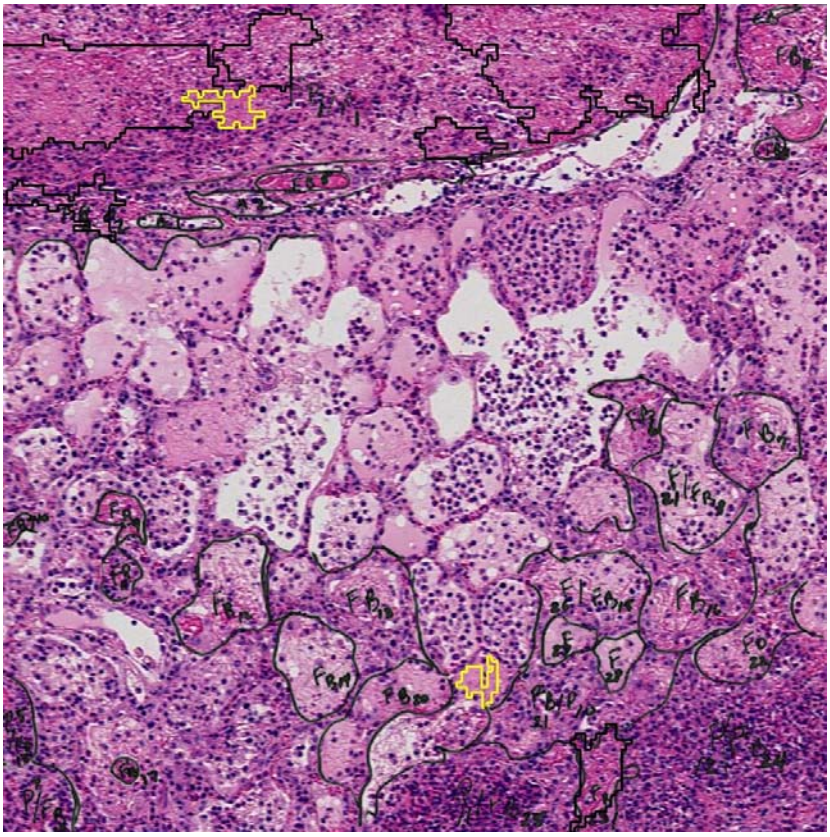


Figure 5.13: Composite image comparing regions of fibrin and necrosis identified in the manual and automated analyses

Table 5.3: Detection of fibrin

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
FB1	Overlaps detected areas FB1, FB2, FB3, FB4	TP
FB2	Not found	FN
FB3	Not found	FN
FB4	Overlaps with FB5	TP
FB5	Not found	FN
FB6	Not found	FN
FB7	Not found	FN
FB8	Not found	FN
FB9	Not found	FN
FB10	Not found	FN
FB11	Not found	FN
FB12	Not found	FN
FB13	Not found	FN
FB14	Not found	FN
FB15	Not found	FN
FB16	Not found	FN
FB17	Not found	FN
FB18	Not found	FN
FB19	Not found	FN
FB20	Not found	FN
FB21	Not found	FN
FB22	Not found	FN
FB23	Almost exact match with FB6	TP
FB24	Not found	FN
FB25	Not found	FN

Number of areas of fibrin = 25

TP = 3 FN = 22 Sensitivity = 12.0%

FP = 0 Positive Predictive Value = 100.0%

Table 5.4: Detection of necrosis

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
None identified	N1	FP
None identified	N2	FP

Number of areas of necrosis = 0
TP = 0 FN = 0 Sensitivity = undefined
FP = 2 Positive Predictive Value = 0.0%

5.2.2 Image Region [2001, 30001] of Slide D03-11282

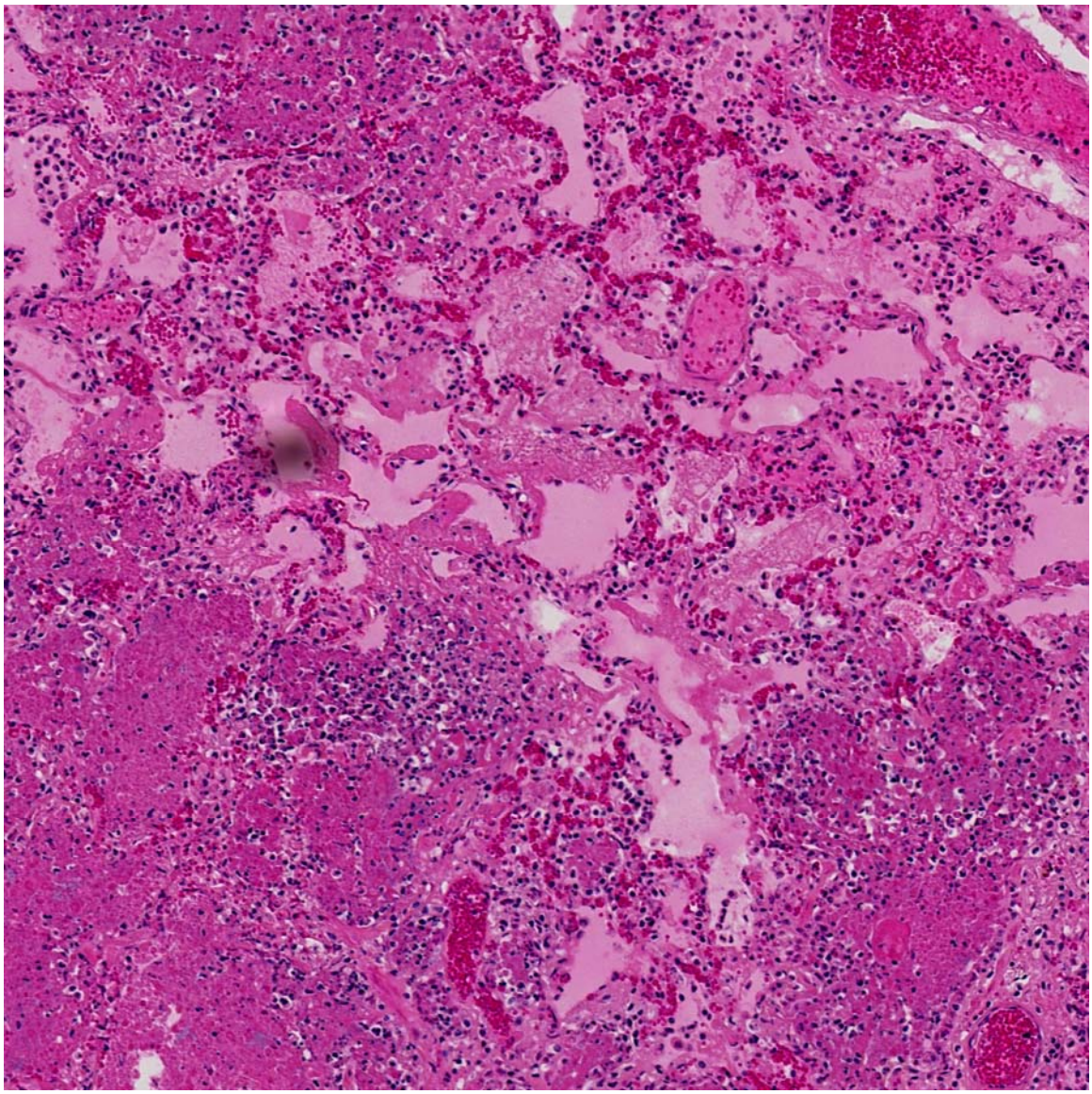


Figure 5.14: Region [2001, 30001] of slide D03-11282

5.2.2.1 Automated Analysis Results

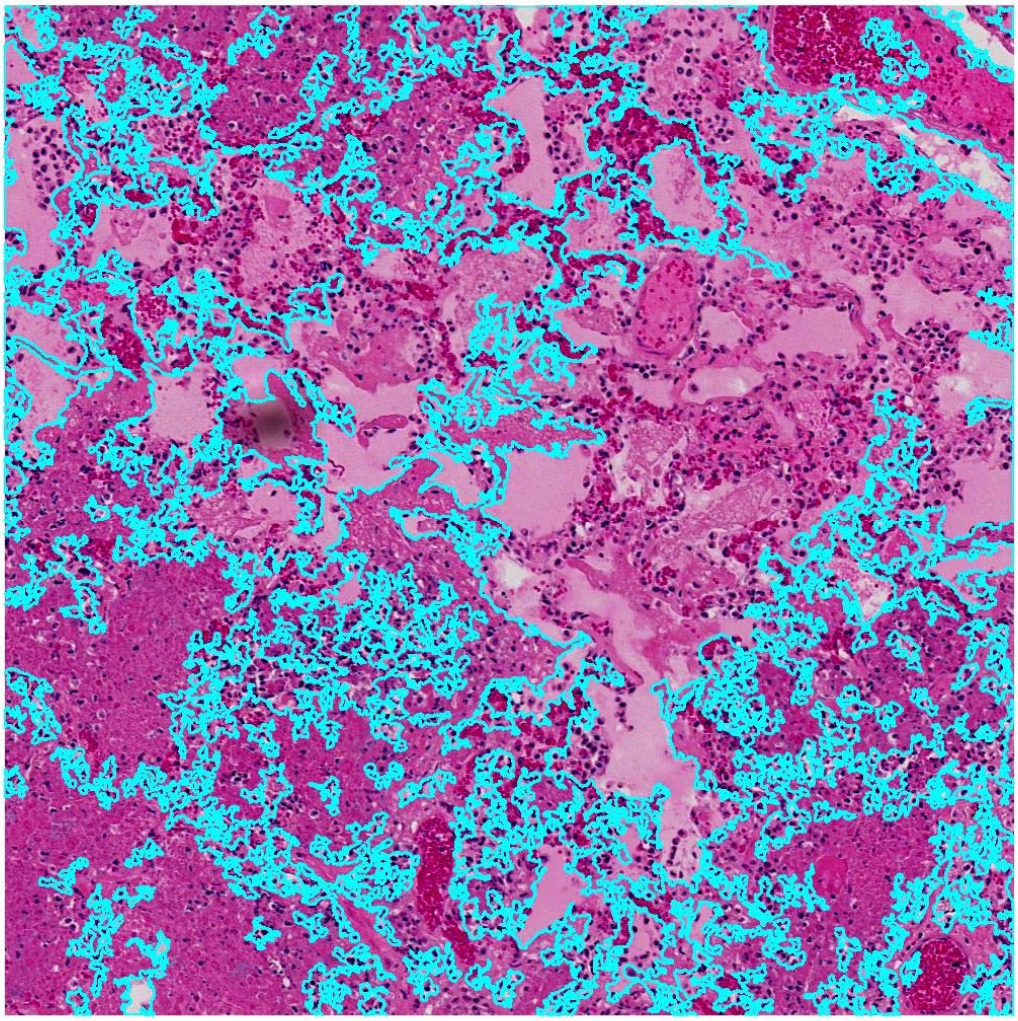


Figure 5.15: All alveoli (air-filled and fluid-filled)

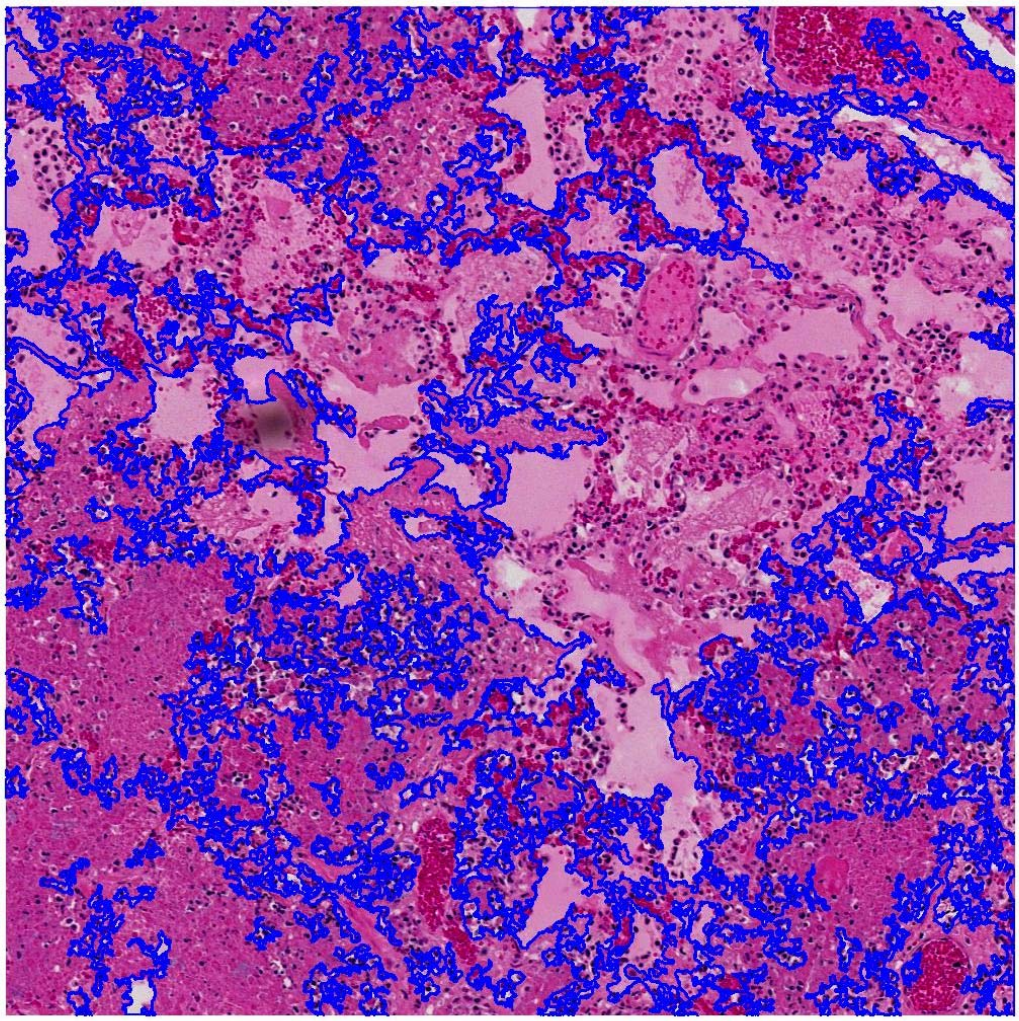


Figure 5.16: Fluid-filled alveoli

```
**********  Image Analysis Report **************************
Image : D0311282_2001x30001
Processing start time : 16-May-2006 15:15:43

Alveoli and Alveolar Contents:
Number of alveoli in image : 125
Percentage of alveoli which are considered fluid-filled (contain more than 33 percent fluid) : 100.00
Total alveolar space in image (pixels) : 2115935
Percentage of total alveolar space which is fluid-filled :  80.84


Areas of Inflammation:
Percentage of tissue area which is considered inflammation :  33.85

********************************************************************************
```

Figure 5.17: Image Analysis Report

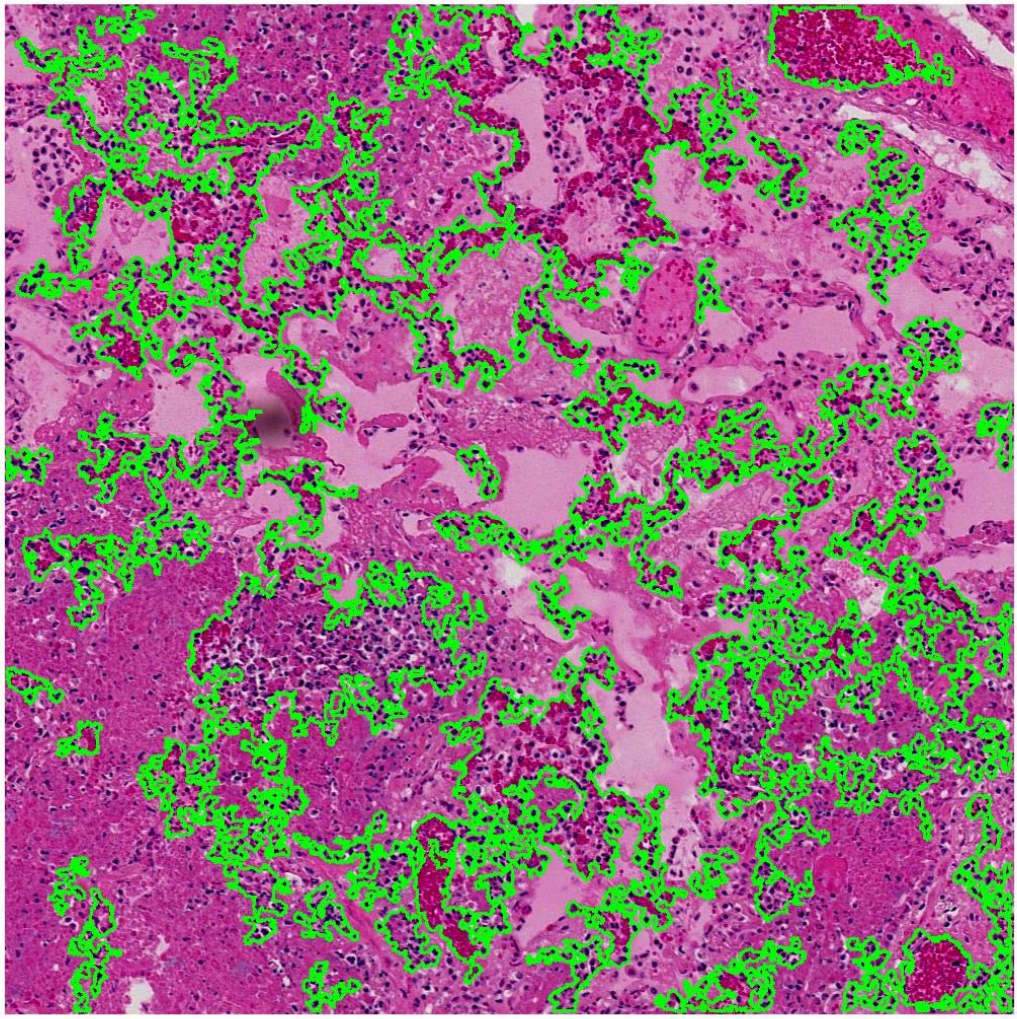


Figure 5.18: Areas of inflammatory cells

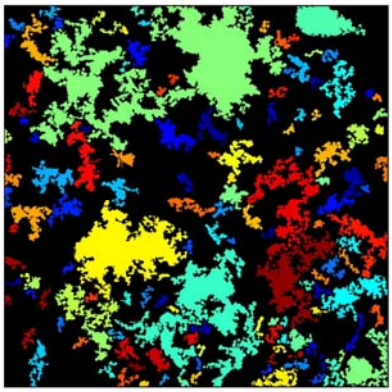


Figure 5.19: Areas of inflammatory cells shown as coloured regions

(Figure 5.19 is provided as an aid in interpreting the regions outlined in the Figure 5.18. Colours assigned to regions in this image are not significant.)

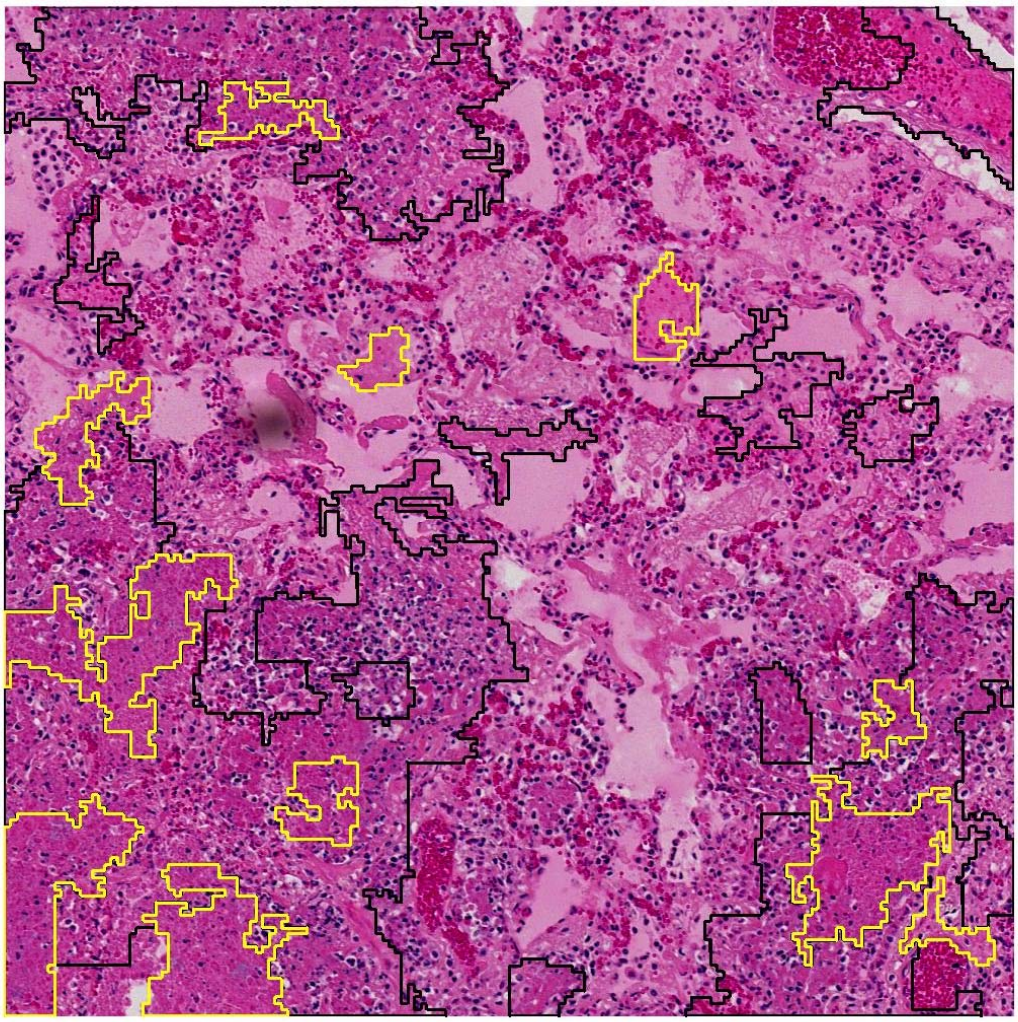


Figure 5.20: Areas of fibrin (outlined in black) and areas of necrosis (outlined in yellow)

```
**********  Fibrin and Necrosis Report **************************
Image : D0311282_2001x30001
Processing start time : 16-May-2006 15:33:58

Percentage of tissue area which is considered fibrin :  30.01

Percentage of tissue area which is considered necrosis :   8.82

*******************************************************************************
```

Figure 5.21: Fibrin and Necrosis Report

5.2.2.2 Manual Analysis Results

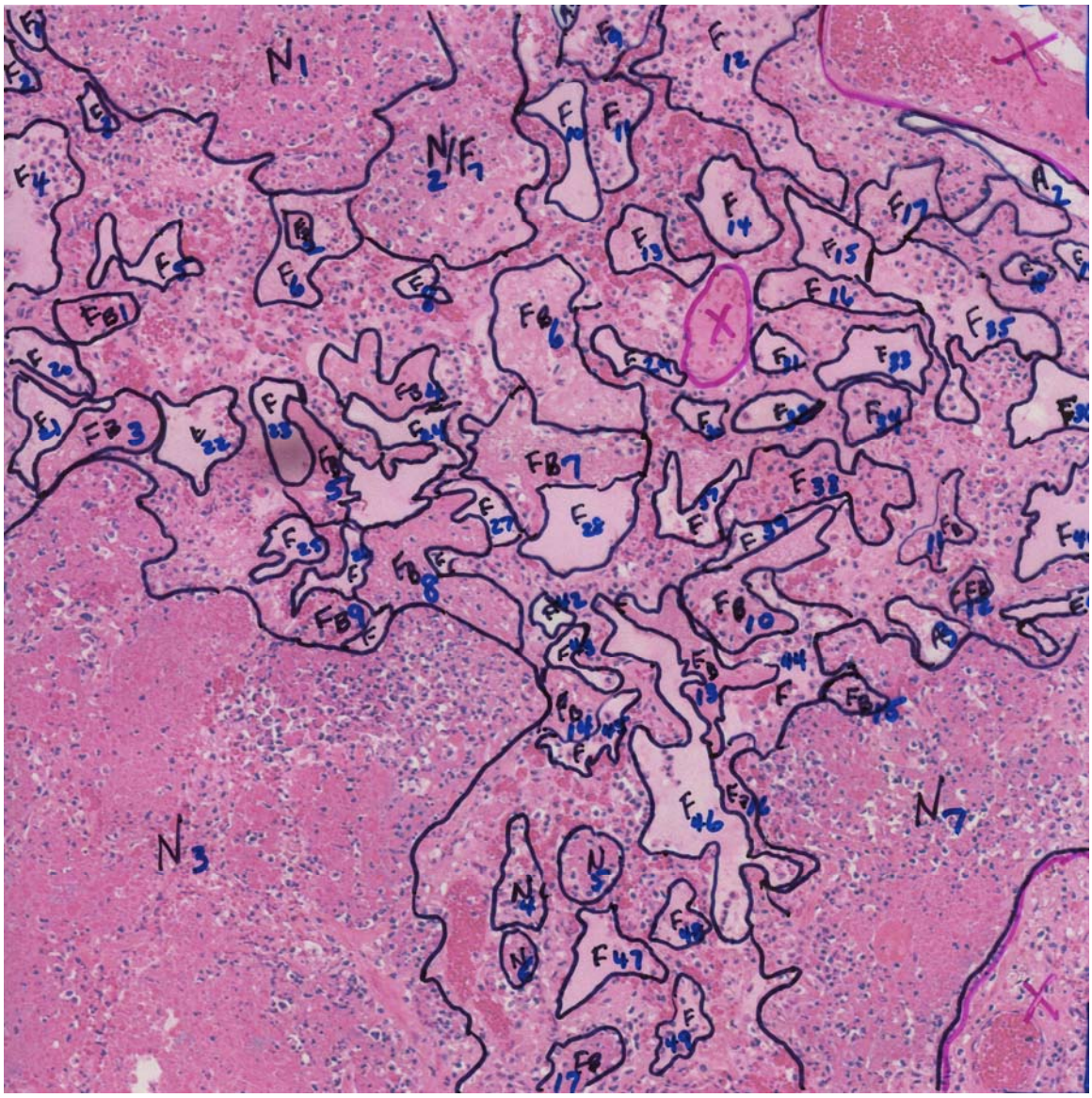


Figure 5.22: Expert manual analysis of region [2001, 30001] of slide D03-11282

Legend:	
A	Air-filled alveolus
F	Fluid-filled alveolus
P	Area of inflammation
FB	Fibrin
N	Necrosis
X	Area of image which is not to be analyzed

Note: In addition to the areas marked in the image above, all background tissue was identified as a combination of fluid and necrosis in the expert manual analysis.

5.2.2.3 Comparison of Results

The image below (Figure 5.23) shows all air-filled and fluid-filled alveoli identified by the manual analysis overlaid on the fluid-filled alveoli image produced by the software. In addition to the specific alveoli marked, the expert manual analysis of this image identified all background tissue as containing fluid and necrosis. A detailed comparison of the manual and automated results for alveoli detection is included in Table 5.5.

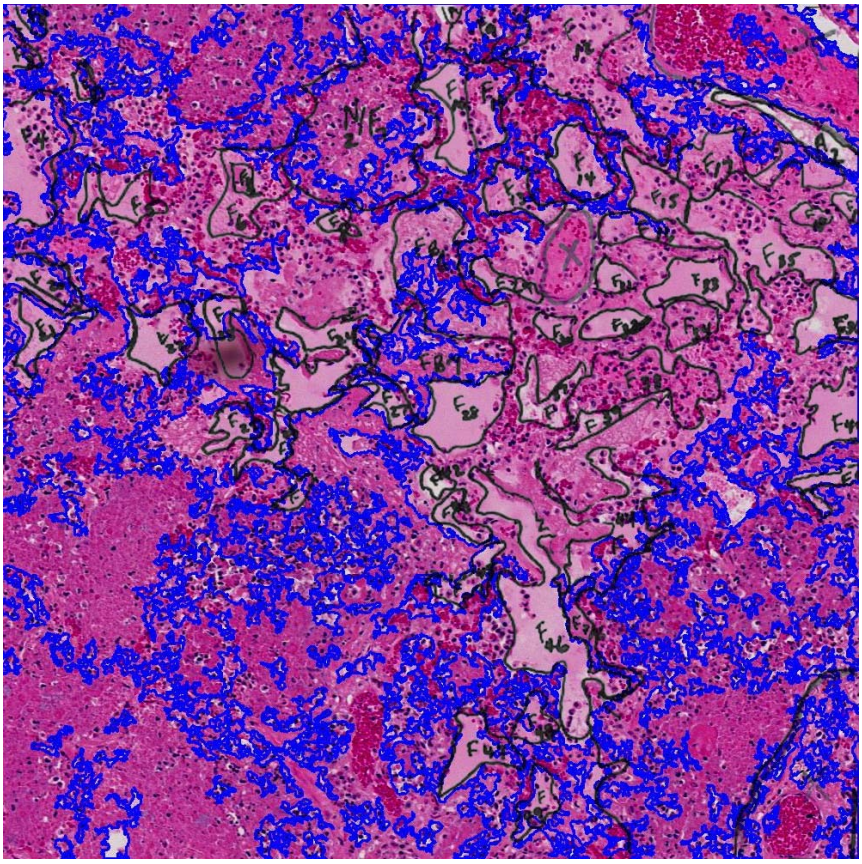


Figure 5.23: Composite image comparing all alveoli detected by the manual and automated analyses

Table 5.5: Detection of all alveoli

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
A1	Found. Identified as part of a larger alveolus, marked as fluid-filled.	TP
A2	Found. Fair match on boundaries.	TP
A3	Found. Fair match on boundaries.	TP
	All air-filled alveoli (A1 – A3) identified as fluid-filled.	FP x 3

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
F1	Found. Excellent match on boundaries	TP
F2	Found as part of a larger fluid-filled alveolus.	TP
F3, F4	Found. Identified as parts of a larger fluid-filled alveolus.	TP x 2
F5, F6, F8, F24	Found. Identified as parts of a larger fluid-filled alveolus.	TP x 4
F7	Partially found as several smaller fluid-filled alveoli. Most of area not identified.	FN
F9	Found as part of larger fluid-filled alveolus	TP
F10, F11	Found. Identified as parts of a larger fluid-filled alveolus. Good match on boundaries (when two areas combined).	TP x 2
F12, F13, F15, F16, F17, F18, F19, F25, F26, F28, F29, F30, F31, F32, F33, F34, F35, F36, F37, F38, F39, F40, F42, F43, F44, F45, F46, F48	All identified as part of one large fluid-filled alveolus (This is consistent with all background areas being identified as fluid and necrosis)	TP x 28
F14	Found. Excellent match on boundaries	TP
F20	Found. Excellent match on boundaries	TP
F21	Found. Excellent match on boundaries	TP
F22	Found. Good match on boundaries.	TP
F23	Partially found. Partially overlaps with larger fluid-filled alveolus.	TP
F27	Found. Good match on boundaries.	TP
F41	Found. Excellent match on boundaries.	TP
F47	Found as part of a slightly larger alveolus	TP
F49	Found. Excellent match on boundaries.	TP
	Small fluid-filled areas identified within large area N1	FP
	Small fluid-filled areas identified within large area N3	FP
	Small fluid-filled areas identified within large area N7	FP
	Small fluid-filled areas identified within area marked “not to be analyzed”	FP
	Areas N4, N5 identified as fluid-filled alveoli	FP x 2
	Small fluid-filled areas identified within areas of fibrin FB1, FB5, FB6, FB7, FB9, FB11, FB15	FP x 7
	Large fluid-filled areas overlap with smaller alveoli identified by Dr. Wobeser, plus cover some of the background area which is identified as fluid / necrosis.	TP x ?

<u>All Alveoli</u>

Number of alveoli (air-filled and fluid-filled) = 52 (specifically marked, plus all background was identified as fluid and necrosis)

TP = 51+++ FN = 1 Sensitivity = 98.1%

FP = 13 Positive Predictive Value = 79.7%

Adjusted FP = 12 Adjusted Positive Predictive Value = 81.0%

Fluid-filled Alveoli

Number of fluid-filled alveoli = 49 (specifically marked, plus all background was identified as fluid and necrosis)

TP = 48+++ FN = 1 Sensitivity = 98.0%

Within set of all objects: FP = 16 Positive Predictive Value = 75.0%

Adjusted FP = 15 Adjusted Positive Predictive Value = 76.2%

Within set of alveoli: FP = 3 Positive Predictive Value = 94.1%

TN = 0 Specificity = 0.0%

For this particular image, no areas of inflammatory cells were identified in the manual analysis. However, numerous areas of inflammatory cells were identified by the software. A detailed comparison of these results is included in Table 5.6.

Table 5.6: Detection of areas of inflammatory cells

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
	Areas of inflammation identified within regions N1, N2, N3, N4, N7	FP x 5
	Areas of inflammation identified within two areas marked as "not to be analyzed"	FP x 2
	Areas of inflammation identified in fluid-filled alveoli F6, F9, F10, F11, F12, F13, F35, F38, F44	FP x 9
	Areas of inflammation identified within areas of fibrin FB5, FB6, FB9, FB15	FP x 4
	Areas of inflammation identified within background, which was identified as fluid / necrosis	FP x ?

Number of areas of inflammatory cells = 0

TP = 0 FN = 0 Sensitivity = undefined

FP = 20+ Positive Predictive Value = 0.0%

Adjusted FP = 18+ Adjusted Positive Predictive Value = 0.0%

The image below (Figure 5.24) shows all areas of fibrin and necrosis identified by the expert manual analysis overlaid on the highlighted fibrin and necrosis image produced by the software. In addition to the specific areas of necrosis marked, the manual analysis also identified all background tissue as containing a combination of fluid and necrosis. A detailed comparison of fibrin detection results is included in Table 5.7, and of necrosis detection results in Table 5.8.

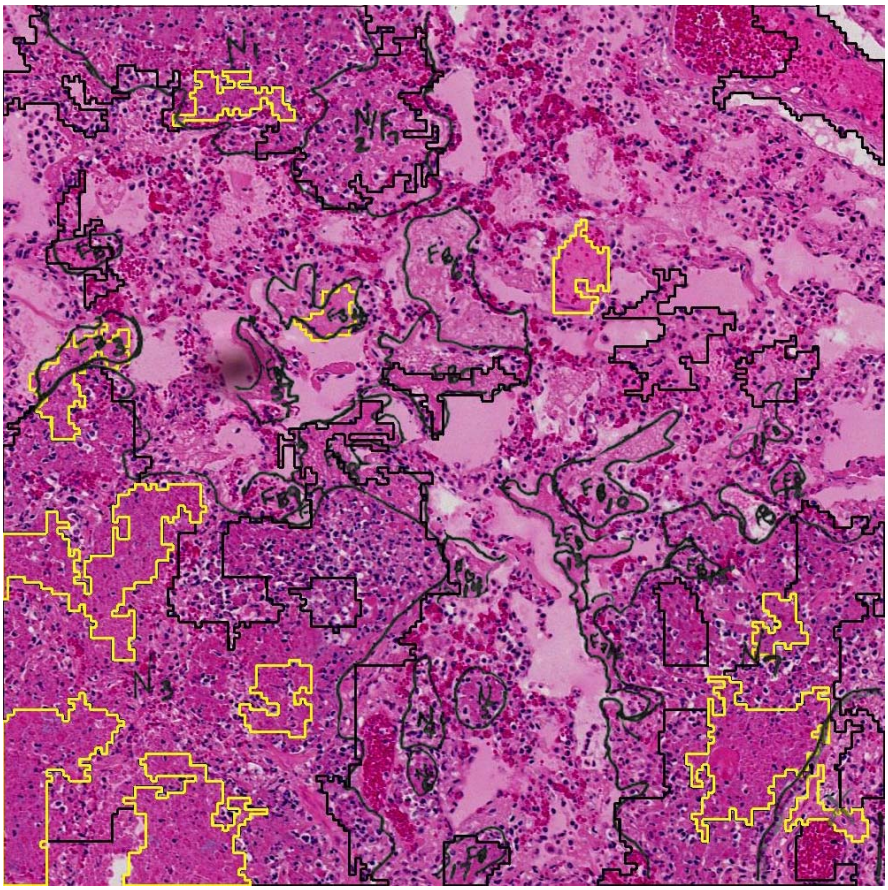


Figure 5.24: Composite image comparing regions of fibrin and necrosis detected by the manual and automated analyses

Table 5.7: Detection of fibrin

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
FB1	Overlaps with FB3.	TP
FB2	Not found.	FN
FB3	Identified as necrosis.	FN
FB4	Identified as necrosis.	FN
FB5	Not found.	FN
FB6	Not found. Partially identified as fluid-filled alveoli.	FN
FB7	Partially identified as region FB4.	TP
FB8	Overlaps with larger area identified as FB8.	TP
FB9	Not found.	FN
FB10	Not found.	FN
FB11	Partially overlapped with region identified as FB6	TP

Region identified by Dr. Wobeser	*Corresponding region identified by programs*	*Count*
FB12	Not found.	FN
FB13	Not found.	FN
FB14	Not found.	FN
FB15	Not found.	FN
FB16	Not found.	FN
FB17	Found as region FB11. Excellent match on boundaries.	TP
	FB1 should be necrosis and necrosis/fluid-filled alveoli.	FP
	FB2 in region identified as “not to be analyzed”	FP
	FB5 overlaps with areas identified as fluid-filled alveoli	FP
	FB6 in background area identified as fluid and necrosis	FP
	FB7, FB8 both in large area identified as necrosis	FP x 2
	FB9, FB10, FB12 all in large area identified as necrosis	FP x 3

Number of areas of fibrin = 17
TP = 5 FN = 12 Sensitivity = 29.4%
FP = 8 Positive Predictive Value = 38.5%
Adjusted FP = 7 Adjusted Positive Predictive Value = 41.7%

Table 5.8: Detection of necrosis

Region identified by Dr. Wobeser	*Corresponding region identified by programs*	*Count*
N1	Partially identified as region N1. Most of area identified as Fibrin.	TP
N2	Not found. Identified as fibrin.	FN
N3	Areas of region identified as regions N3, N5, N6, N7, N8	TP
N4	Not found.	FN
N5	Not found.	FN
N6	Not found.	FN
N7	Areas of region identified as N9, N10	TP
	N2 overlaps with region FB4	FP
	N4 located in region identified as “not to be analyzed”	FP

Number of areas of necrosis = 7
TP = 3 FN = 4 Sensitivity = 42.9%
FP = 2 Positive Predictive Value = 60.0%

5.2.3 Image Region [8001, 28001] of Slide D03-51216

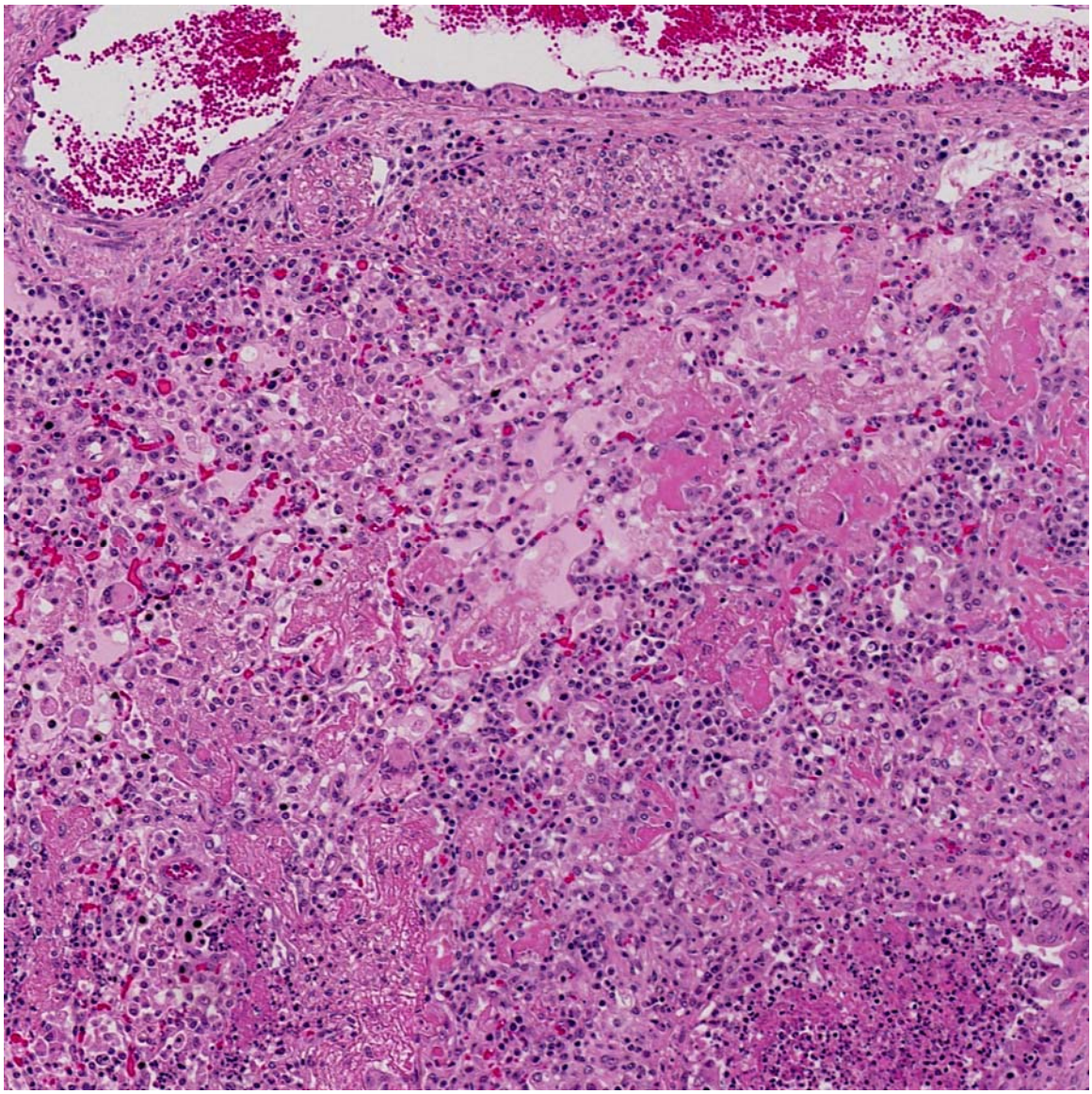


Figure 5.25: Region [8001, 28001] of slide D03-51216

5.2.3.1 Automated Analysis Results

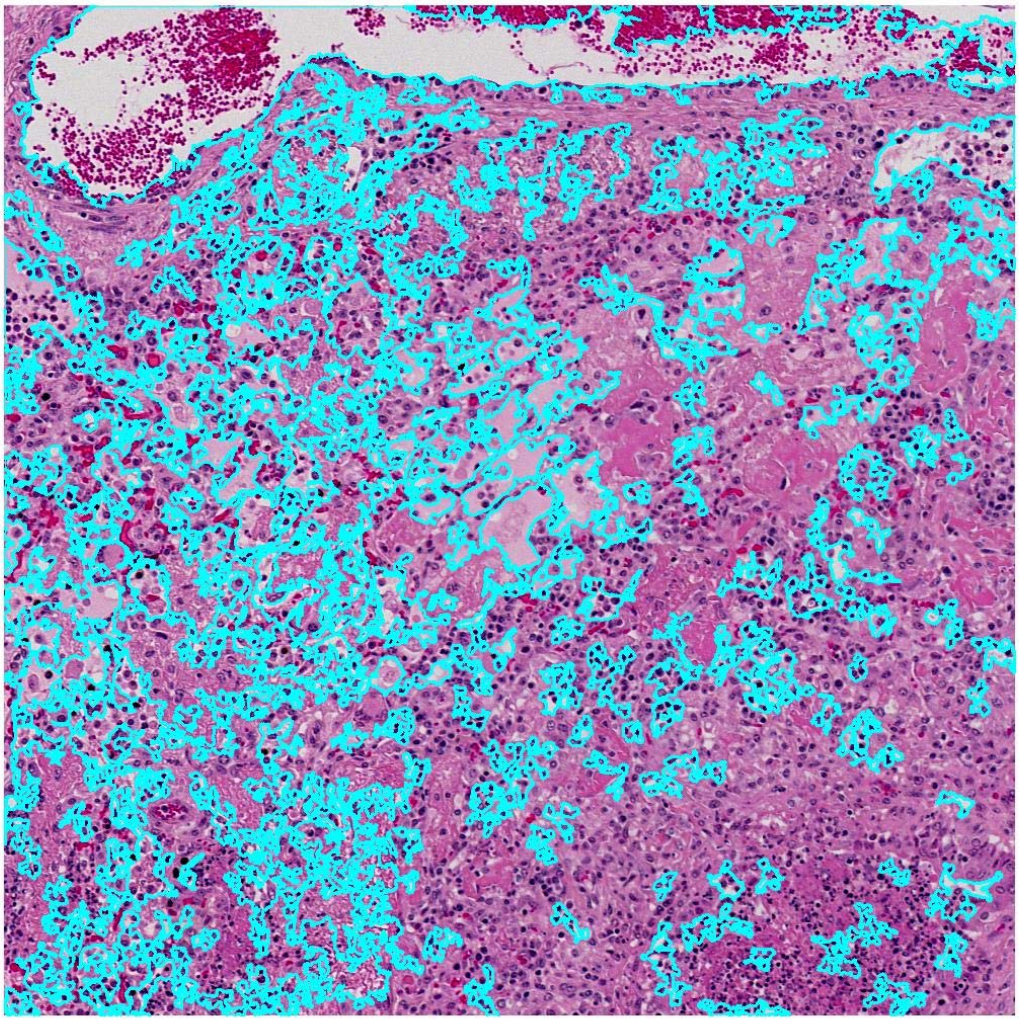


Figure 5.26: All alveoli (air-filled and fluid-filled)

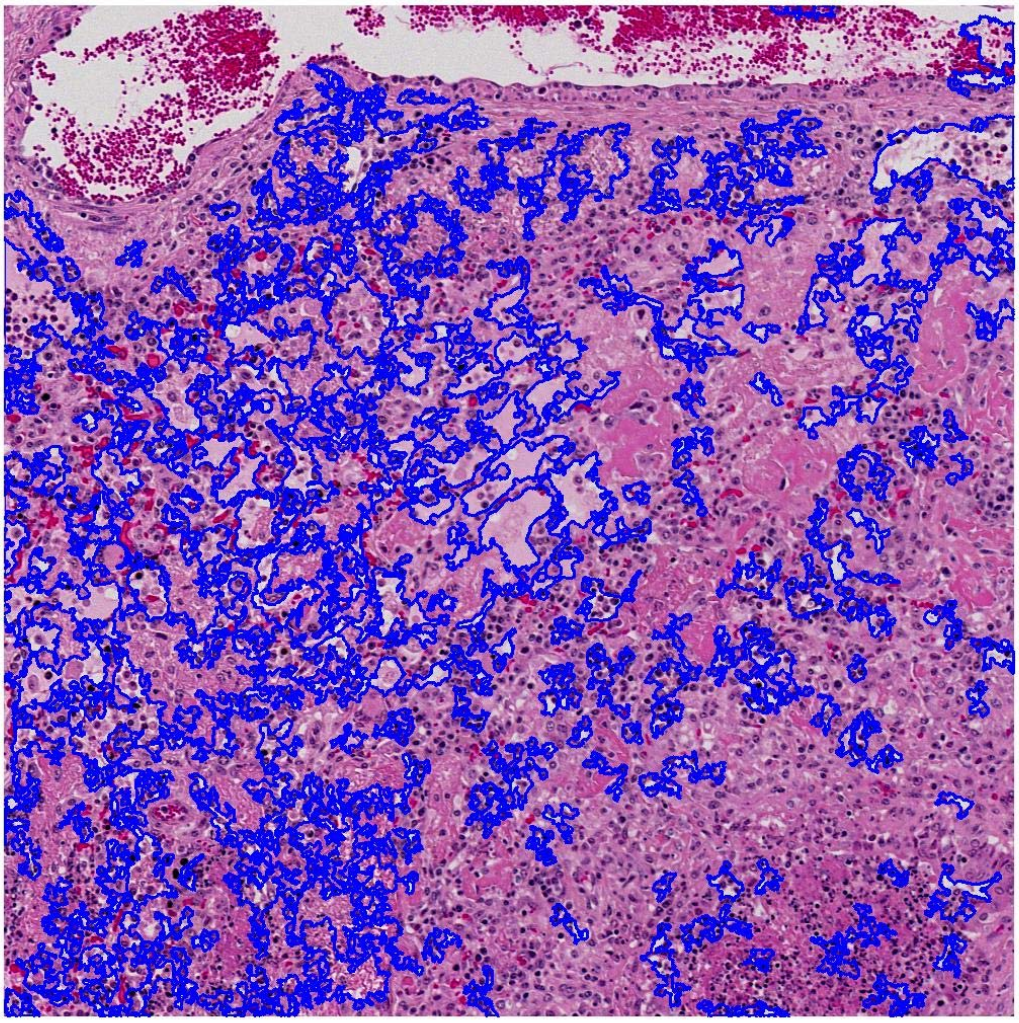


Figure 5.27: Fluid-filled alveoli

********** Image Analysis Report **********
Image : D0311282_2001x30001
Processing start time : 16-May-2006 15:15:43

Alveoli and Alveolar Contents:
Number of alveoli in image : 125
Percentage of alveoli which are considered fluid-filled (contain more than 33 percent fluid) : 100.00
Total alveolar space in image (pixels) : 2115935
Percentage of total alveolar space which is fluid-filled : 80.84

Areas of Inflammation:
Percentage of tissue area which is considered inflammation : 33.85

Figure 5.28: Image Analysis Report

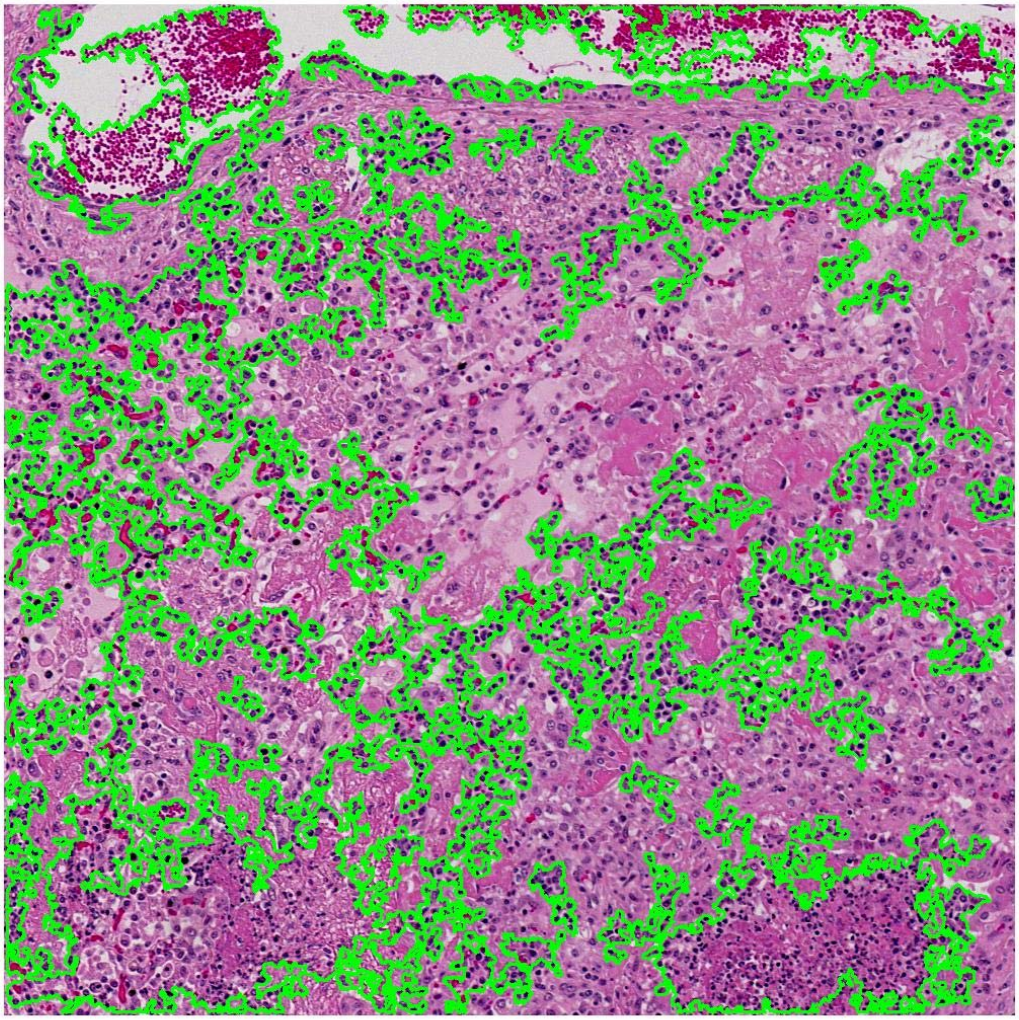


Figure 5.29: Areas of inflammatory cells

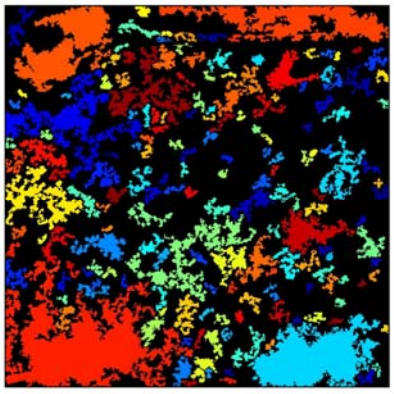


Figure 5.30: Areas of inflammatory cells shown as coloured regions

(Figure 5.30 is provided as an aid in interpreting the regions outlined in the Figure 5.29. Colours assigned to regions in this image are not significant.)

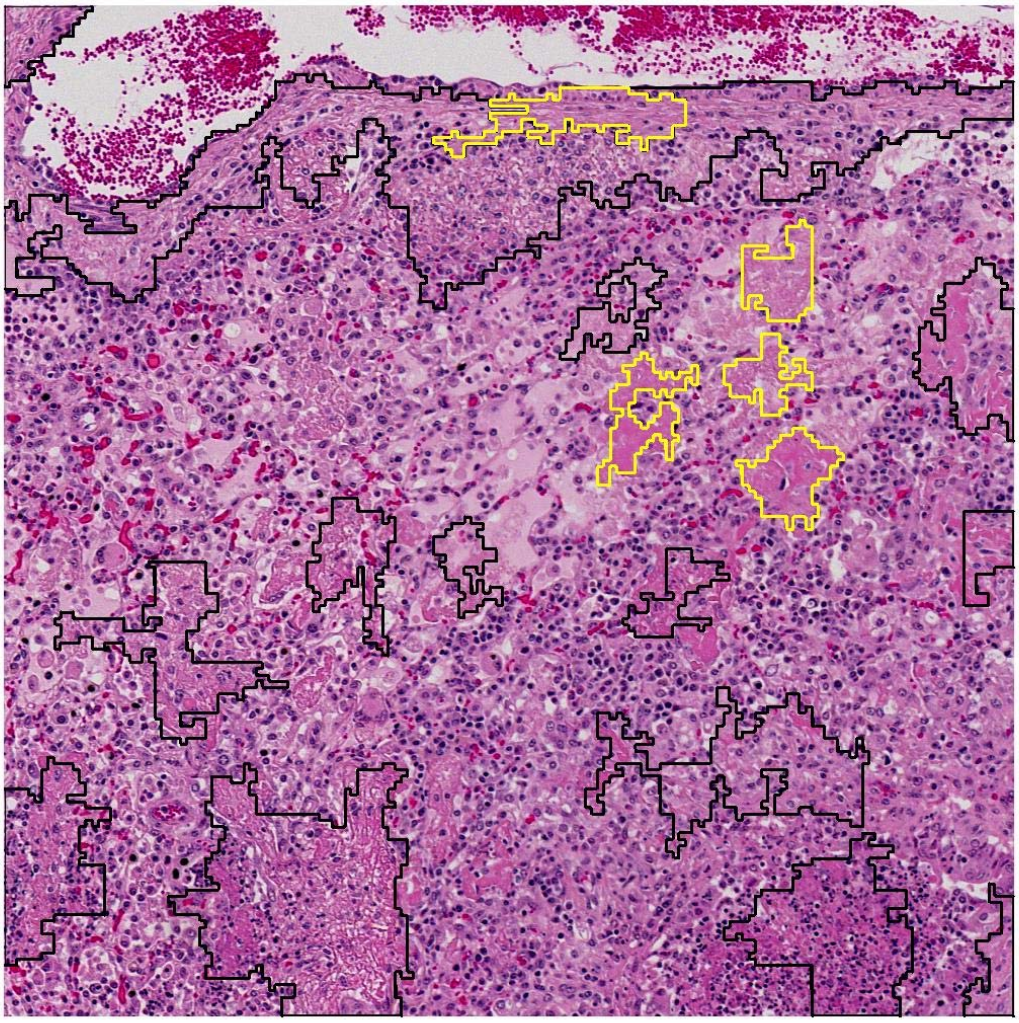


Figure 5.31: Areas of fibrin (outlined in black) and areas of necrosis (outlined in yellow)

```
**********  Fibrin and Necrosis Report **************************
Image : D0351216_8001x28001
Processing start time : 18-May-2006 13:15:11

Percentage of tissue area which is considered fibrin :  22.70

Percentage of tissue area which is considered necrosis :   2.64

*******************************************************************************
```

Figure 5.32: Fibrin and Necrosis Report

5.2.3.2 Manual Analysis Results

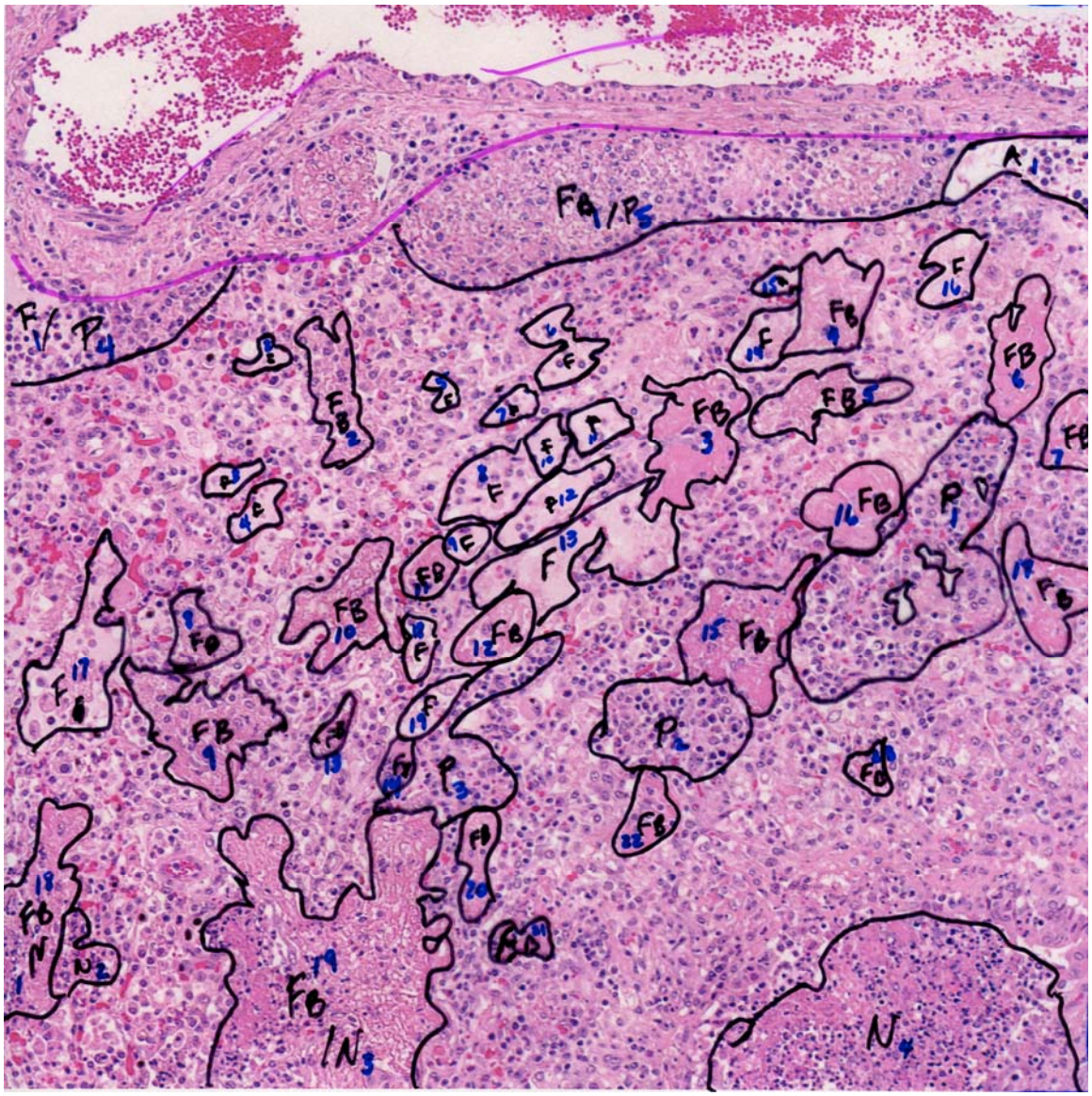


Figure 5.33: Expert manual analysis of region [8001, 28001] of slide D03-51216

Legend:	
A	Air-filled alveolus
F	Fluid-filled alveolus
P	Area of inflammation
FB	Fibrin
N	Necrosis
X	Area of image which is not to be analyzed

Note: In addition to the areas marked in the image above, all background tissue was identified as a combination of fluid and inflammation in the expert manual analysis.

5.2.3.3 Comparison of Results

The image below (Figure 5.34) shows all air-filled and fluid-filled alveoli identified by the manual analysis overlaid on the fluid-filled alveoli image produced by the software. For this image, the software identified all alveoli in this image as fluid-filled, so this one image shows the complete set of alveoli detected by the programs. In addition to all alveoli specifically marked, the expert manual analysis also identified all background tissue as containing a combination of fluid and inflammation. A detailed comparison of the manual and automated results for alveoli detection is included in Table 5.9.

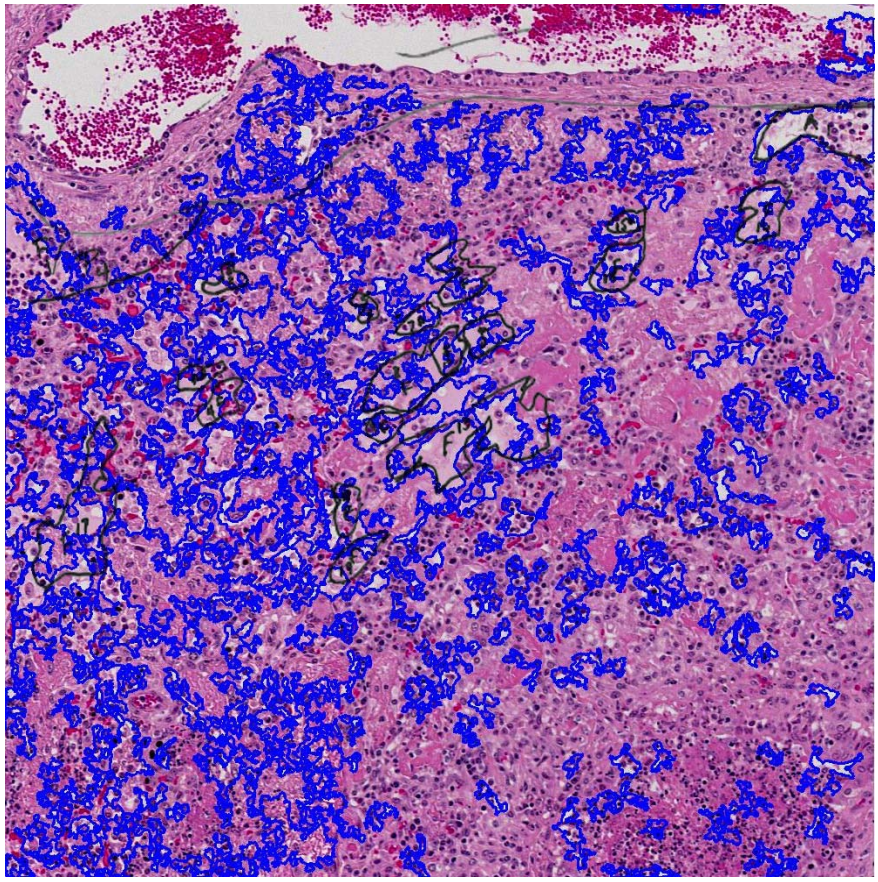


Figure 5.34: Composite image comparing all alveoli detected by the manual and automated analyses

Table 5.9: Detection of all alveoli

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
A1	Found. Excellent match on boundaries.	TP
	A1 identified as fluid-filled.	FP
	Large alveolus marked in area identified as “not to be analyzed”	FP
F1	Found. Overlaps with a few smaller fluid-filled alveoli.	TP
F2	Found. Overlaps with larger fluid-filled alveolus.	TP
F3, F4	Found as part of a larger alveolus	TP x 2
F5, F7, F10, F11, F12, F15, F16	Found. Good match on boundaries	TP x 7
F6	Found. Overlaps with two smaller alveoli.	TP
F8	Found. Overlaps with two smaller alveoli.	TP
F9	Found. Fair match on boundaries.	TP
F13	Found. Overlaps with three smaller alveoli.	TP
F14	Found. Overlaps with two smaller alveoli.	TP
F17	Found. Overlaps with a few smaller alveoli.	TP
F18	Found. Overlaps with larger alveolus.	TP
F19	Found. Overlaps with two smaller alveoli.	TP
All unmarked tissue areas identified as fluid and inflammation	Numerous additional fluid-filled alveoli marked.	TP x many
	Small fluid-filled areas marked within areas of fibrin FB1, FB2, FB9, FB10, FB12, FB13, FB18, FB19, FB22	FP x 9
	Small fluid-filled areas marked within area of necrosis N4	FP
	Small fluid-filled areas marked within areas of inflammation P1, P2, P3	FP x 3
	A few fluid-filled alveoli marked in area identified as “not to be analyzed”	FP x 3

<u>All Alveoli</u>

Number of alveoli (air-filled and fluid-filled) = 20 (specifically marked, plus all background was identified as fluid and inflammation)

TP = 20+++ FN = 0 Sensitivity = 100.0%

FP = 16 Positive Predictive Value = 55.6%

Adjusted FP = 13 Adjusted Positive Predictive Value = 60.6%

Fluid-Filled Alveoli

Number of fluid-filled alveoli = 19 (specifically marked, plus all background was identified as fluid and inflammation)

TP = 19 +++ FN = 0	Sensitivity = 100.0%	
Within set of all objects:	FP = 17	Positive Predictive Value = 52.8%
	Adjusted FP = 14	Adjusted Positive Predictive Value = 57.6%
Within set of alveoli:	FP = 1	Positive Predictive Value = 95.0%
	TN = 0	Specificity = 0.0%

The image below (Figure 5.35) shows all regions of inflammatory cells identified in the manual analysis overlaid over the corresponding regions located by the software. In addition to the specifically marked areas, the manual analysis also identified all background tissue as containing a combination of fluid and inflammation. A detailed comparison of the manual and automated results is included in Table 5.10.

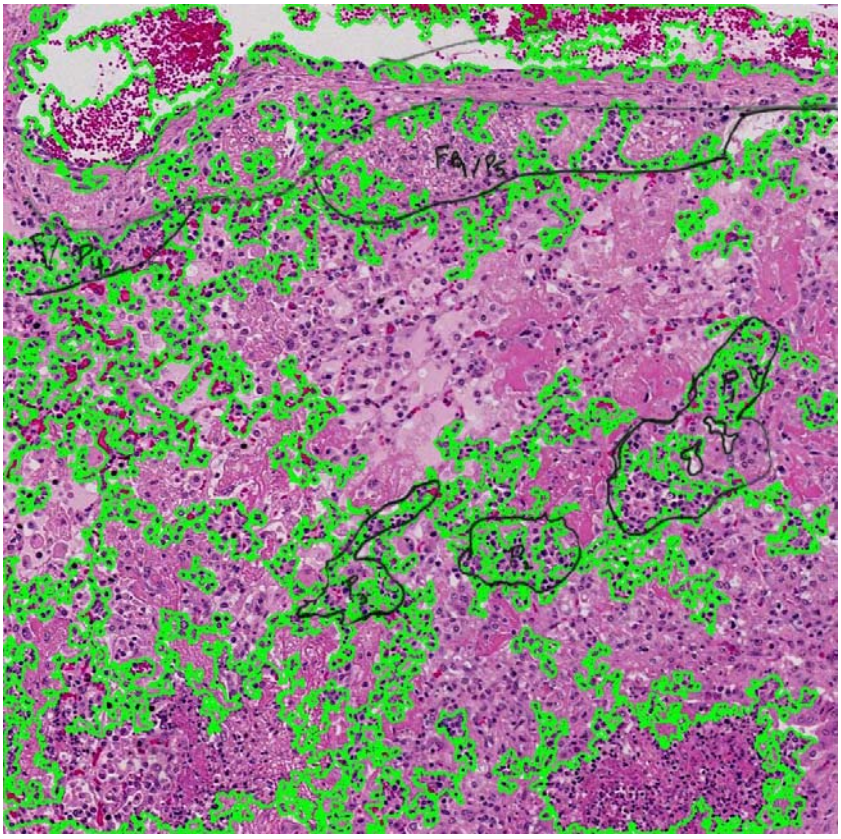


Figure 5.35: Composite image comparing areas of inflammation detected by the manual and automated analyses

Table 5.10: Detection of areas of inflammatory cells

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
P1	Found. Overlaps with a few smaller areas of inflammatory cells.	TP
P2	Found. Overlaps with a few smaller areas of inflammatory cells.	TP
P3	Found. Overlaps with a few smaller areas of inflammatory cells.	TP
P4	Found. Overlaps with a few smaller areas of inflammatory cells.	TP
P5	Found. Overlaps with a few smaller areas of inflammatory cells.	TP
All unmarked tissue areas identified as fluid and inflammation	Numerous additional areas of inflammatory cells marked.	TP x many
	Areas of inflammatory cells identified within alveolus A1	FP
	Areas of inflammatory cells identified within areas of fibrin FB14, FB15, FB18, FB19	FP x 4
	Areas of inflammatory cells identified within areas of necrosis N4	FP
	A few areas of inflammatory cells marked in area identified as “not to be analyzed”	FP x 3

Areas of inflammatory cells = 5 (specifically marked, plus all background was identified as fluid and inflammatory cells)
TP = 5+++ FN = 0 Sensitivity = 100.0%
FP = 9 Positive Predictive Value = 35.7%
Adjusted FP = 6 Adjusted Positive Predictive Value = 45.5%

The image below (Figure 5.36) shows all regions of fibrin and necrosis identified by the manual analysis overlaid on the corresponding regions identified by the software. A detailed comparison of these manual and automated analysis results is included in Table 5.11 for fibrin, and Table 5.12 for necrosis.

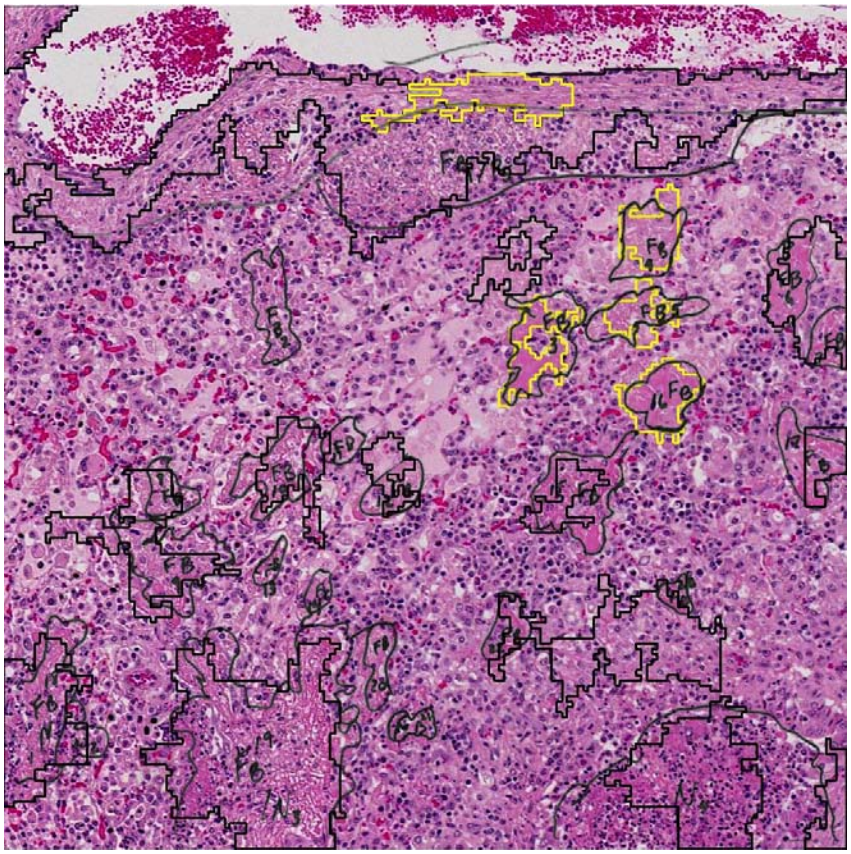


Figure 5.36: Composite image comparing regions of fibrin and necrosis identified by the manual and automated analyses

Table 5.11: Detection of fibrin

Region identified by Dr. Wobeser	***Corresponding region identified by programs***	***Count***
FB1	Found. Overlaps with region FB1.	TP
FB2	Not found.	FN
FB3	Not found. Identified as necrosis area N4	FN
FB4	Not found. Identified as necrosis area N2	FN
FB5	Not found. Identified as necrosis area N3	FN
FB6, FB7	Found as part of one larger area of fibrin, FB3	TP x 2
FB8, FB9	Found as part of one larger area of fibrin, FB4	TP x 2
FB10	Found. Overlaps with region FB5	TP
FB11	Partially overlaps with region FB5. Most of area not identified.	FN
FB12	Found. Overlaps with area FB6.	TP
FB13, FB14	Not found	FN x 2
FB15	Found. Overlaps with region FB7.	TP

Region identified by Dr. Wobeser	*Corresponding region identified by programs*	*Count*
FB16	Not found. Identified as necrosis area N5.	FN
FB17	Found. Overlaps with region FB8.	TP
FB18	Found as region FB9. Good match on boundaries.	TP
FB19	Found as region FB19. Good match on boundaries.	TP
FB20, FB21	Not found.	FN x 2
FB22, FB23	Found as part of larger fibrin area FB11	TP x 2
	FB2. No corresponding region identified.	FP
	FB12, FB13. Corresponds to region N4 identified as necrosis.	FP x 2

Number of areas of fibrin = 23
TP = 13 FN = 10 Sensitivity = 56.5%
FP = 3 Positive Predictive Value = 81.3%

Table 5.12: Detection of necrosis

Region identified by Dr. Wobeser	*Corresponding region identified by programs*	*Count*
N1, N2	Not found. Marked as fibrin area FB9	FN x 2
N3	Not found. Marked as fibrin area FB10	FN
N4	Not found. Marked as fibrin areas FB12, FB13.	FN
	N1. Marked in area identified as “not to be analyzed”	FP
	N2. Corresponds to fibrin region FB4. Good match on boundaries.	FP
	N3. Corresponds to fibrin region FB5. Good match on boundaries.	FP
	N4. Corresponds to fibrin region FB3. Good match on boundaries.	FP
	N5. Corresponds to fibrin region FB16. Excellent match on boundaries.	FP

Number of areas of necrosis = 4
TP = 0 FN = 4 Sensitivity = 0.0%
FP = 5 Positive Predictive Value = 0.0%
Adjusted FP = 4 Adjusted Positive Predictive Value = 0.0%

5.3 Results of Alveoli Detection

5.3.1 Measurement of Results

Detection results were scored for two groups of alveoli. The "All Alveoli" results (Table 5.13) look at the success of detecting any alveoli, whether air-filled or fluid-filled. For this count, similar to those of inflammation, fibrin and necrosis, no specificity value is calculated since the number of possible false positives is unlimited. The "All Alveoli" count does not examine identification of the alveolar contents. If an alveolus was found, it is counted as a true positive, regardless of how its contents were identified. The second set of alveoli results is for fluid-filled alveoli only. For these results, a true positive is scored only if an alveolus is found and also marked as fluid-filled. False negatives include fluid-filled alveoli which were not detected at all, as well as those that were found but were incorrectly identified as air-filled. For the fluid-filled alveoli count, false positives and positive predictive value are counted in two different ways. First, the number of false positives within the full set of objects is counted (Table 5.14). This value would include regions of fibrin which were incorrectly identified as fluid-filled alveoli, for example, as well as any air-filled alveoli which were wrongly thought to contain fluid. Similar to the results tallies for all other objects, no true negative value can be counted in this case, so specificity is not calculated. The second set of results for fluid-filled alveoli examines scores within the set of alveoli only (Table 5.15). In this case, the total number of objects is the number of alveoli in the image, the false positive count is only the number of air-filled alveoli which were identified as containing fluid, and the true negative value is the number of air-filled alveoli not identified as fluid-filled. Since the total number of objects is known, it is possible to calculate specificity in this case, and so this value is included, along with the number of false positives and positive predictive value.

In 23 of the 50 images, the entire background tissue was nonspecifically labeled as fluid in the manual analysis, in addition to specific air-filled and fluid-filled alveoli being marked. Counting the number of actual true positive regions in these images is difficult, since there is no set number of areas. For these images, the true positive count was limited to the number of specific alveoli marked in the manual analysis, even though this number is much lower than the actual value should be. In these cases, the number listed in the "Number of Areas" column of the results chart includes a "+++" symbol, indicating that many more fluid-filled areas existed.

Table 5.13: Summary of alveoli detection results for all images

		All Alveoli									
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*
D03-10538	12001, 20001	49	49	0	100.0%		17	17	74.2%	74.2%	
	12001, 36001	20+++	20+++	0	100.0%		0	0	100.0%	100.0%	
	18001, 10001	3	3	0	100.0%		2	2	60.0%	60.0%	
	18001, 40001	29	29	0	100.0%		2	2	93.5%	93.5%	
	30001, 40001	5	5	0	100.0%		17	17	22.7%	22.7%	
D03-00716	18001, 26001	17+++	17+++	0	100.0%		1	1	94.4%	94.4%	
	22001, 52001	2+++	1+++	1	?	??	0	0	undef.	undef.	
	14001, 24001	51	49	2	96.1%		0	0	100.0%	100.0%	
	8001, 22001	24+++	24+++	0	100.0%		1	0	96.0%	100.0%	
	34001, 28001	17+++	16+++	1	94.1%		0	0	100.0%	100.0%	
D03-45740	8001, 22001	47	46	1	97.9%		4	4	92.0%	92.0%	
	18001, 42001	5+++	5+++	0	100.0%		6+	0	45.5%	100.0%	??
	32001, 20001	35	35	0	100.0%		34+	34+	50.7%	50.7%	??
	14001, 20001	32	32	0	100.0%		6+	6+	84.2%	84.2%	??
	6001, 20001	3	3	0	100.0%		9+	6+	25.0%	33.3%	??
D03-11282	4001, 24001	51	51	0	100.0%		2	0	96.2%	100.0%	
	2001, 30001	52+++	51+++	1	98.1%		13	12	79.7%	81.0%	
	24001, 46001	36	33	3	91.7%		3	2	91.7%	94.3%	
	14001, 22001	81	76	5	93.8%		8	8	90.5%	90.5%	
	26001, 34001	77	69	8	89.6%		11	11	86.3%	86.3%	??
D96-4925	18001, 2001	23	23	0	100.0%		0	0	100.0%	100.0%	
	24001, 26001	16	16	0	100.0%		0	0	100.0%	100.0%	
	12001, 32001	9+++	9+++	0	100.0%		1	0	90.0%	100.0%	
	2001, 16001	47+++	47+++	0	100.0%		1	0	97.9%	100.0%	
	16001, 44001	33+++	33+++	0	100.0%		9	0	78.6%	100.0%	
D03-51216	6001, 20001	88+++	88+++	0	100.0%		13	8	87.1%	91.7%	
	10001, 30001	40	40	0	100.0%		4+	0	90.9%	100.0%	??
	8001, 4001	38	34	4	89.5%		8+	8	81.0%	81.0%	??
	8001, 28001	20+++	20+++	0	100.0%		16	13	55.6%	60.6%	
	4001, 12001	53+++	53+++	0	100.0%		7	7	88.3%	88.3%	
D03-37303	12001, 28001	30+++	26+++	4	86.7%		4	0	86.7%	100.0%	
	24001, 16001	48	48	0	100.0%		50+	50+	49.0%	49.0%	??
	22001, 26001	10+++	10+++	0	100.0%		1	1	90.9%	90.9%	??
	8001, 42001	3+++	3+++	0	100.0%		9+	0	25.0%	100.0%	??
	2001, 42001	1+++	1+++	0	100.0%	??	0	0	100.0%	100.0%	
D03-36878	1, 22001	74	66	8	89.2%		18	18	78.6%	78.6%	
	18001, 14001	86	86	0	100.0%		3	3	96.6%	96.6%	
	18001, 8001	112	109	3	97.3%		0	0	100.0%	100.0%	
	22001, 12001	35	35	0	100.0%		2	1	94.6%	97.2%	
	20001, 6001	110	107	3	97.3%		6	6	94.7%	94.7%	
N95-4825	20001, 10001	52+++	52+++	0	100.0%		1	1	98.1%	98.1%	??
	10001, 40001	43+++	39+++	4	90.7%		1	0	97.5%	100.0%	??
	22001, 24001	12+++	12+++	0	100.0%		1	1	92.3%	92.3%	
	8001, 12001	19+++	19+++	0	100.0%		2	2	90.5%	90.5%	
	14001, 40001	53+++	51+++	2	96.2%		2+	0	96.4%	100.0%	??
D03-51213	12001, 18001	39	37	2	94.9%		18	10	67.3%	78.7%	
	6001, 12001	78	59	19	75.6%		3	1	95.2%	98.3%	??
	4001, 6001	82+++	77+++	5	93.9%		1	0	98.7%	100.0%	??
	10001, 18001	51	42	9	82.4%		8	5	84.0%	89.4%	
	16001, 16001	112	89	13	87.3%		11	11	89.0%	89.0%	
	AVERAGE				**94.8%**		**6.7**	**5.4**	**83.2%**	**88.2%**	

(For an explanation of rows indicated by a marker in the "*" column, refer to section 5.3.2)

Table 5.14: Summary of fluid-filled alveoli detection results for all images (Part 1)

		Fluid-Filled Alveoli									
							Within set of all objects				
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*
D03-10538	12001, 20001	30	30	0	100.0%		36	36	45.5%	45.5%	
	12001, 36001	8+++	8+++	0	100.0%		11	11	42.1%	42.1%	
	18001, 10001	1	1	0	100.0%		4	4	20.0%	20.0%	
	18001, 40001	17	17	0	100.0%		13	13	56.7%	56.7%	
	30001, 40001	3	3	0	100.0%		19	19	13.6%	13.6%	
D03-00716	18001, 26001	many	many	0	?	??	18	18	?	?	??
	22001, 52001	2+++	1+++	1	?	??	0	0	100.0%	100.0%	
	14001, 24001	31	29	2	93.5%		20	20	59.2%	59.2%	
	8001, 22001	0+++	0+++	0	undef.		24	23	?	?	??
	34001, 28001	1+++	1+++	0	?	??	15	15	?	?	??
D03-45740	8001, 22001	26	26	0	100.0%		25	25	51.0%	51.0%	
	18001, 42001	many	many	0	?	??	6+	0	?	100.0%	??
	32001, 20001	1	1	0	100.0%		34+	34+	2.9%	2.9%	??
	14001, 20001	26	26	0	100.0%		6+	6+	81.3%	81.3%	??
	6001, 20001	3	3	0	100.0%		9+	6+	25.0%	33.3%	??
D03-11282	4001, 24001	49	49	0	100.0%		4	2	92.5%	96.1%	
	2001, 30001	49+++	48+++	1	98.0%		16	15	75.0%	76.2%	
	24001, 46001	0	0	0	undef.		16	16	0.0%	0.0%	
	14001, 22001	14	14	0	100.0%		70	70	16.7%	16.7%	
	26001, 34001	11	5	6	45.5%		12	12	29.4%	29.4%	
D96-4925	18001, 2001	2	2	0	100.0%		21	21	8.7%	8.7%	
	24001, 26001	0	0	0	undef.		16	16	0.0%	0.0%	
	12001, 32001	4+++	2+++	2	50.0%	??	6	6	?	?	??
	2001, 16001	17+++	17+++	0	100.0%		31	31	35.4%	35.4%	
	16001, 44001	20+++	20+++	0	100.0%		22	22	47.6%	47.6%	
D03-51216	6001, 20001	67+++	67+++	0	100.0%		34	34	66.3%	66.3%	
	10001, 30001	13	13	0	100.0%		31+	31+	29.5%	29.5%	??
	8001, 4001	15	14	1	93.3%		31	31	31.1%	31.1%	
	8001, 28001	19+++	19+++	0	100.0%		17	14	52.8%	57.6%	
	4001, 12001	51+++	51+++	0	100.0%		9	9	85.0%	85.0%	
D03-37303	12001, 28001	0+++	0+++	0	undef.		27	27	?	?	??
	24001, 16001	3	3	0	100.0%		50+	50+	5.7%	5.7%	??
	22001, 26001	0+++	0+++	0	undef.		11	11	?	?	??
	8001, 42001	3+++	3+++	0	100.0%		9+	0	?	100.0%	??
	2001, 42001	1+++	1+++	0	100.0%		0	0	100.0%	100.0%	
D03-36878	1, 22001	15	9	6	60.0%		77	77	10.5%	10.5%	
	18001, 14001	34	34	0	100.0%		55	55	38.2%	38.2%	
	18001, 8001	38	36	2	94.7%		74	74	32.7%	32.7%	
	22001, 12001	18	18	0	100.0%		19	19	48.6%	48.6%	
	20001, 6001	17	17	0	100.0%		99	99	14.7%	14.7%	
N95-4825	20001, 10001	11+++	11+++	0	100.0%		42	42	20.8%	20.8%	
	10001, 40001	19+++	16+++	3	84.2%		13	13	55.2%	55.2%	
	22001, 24001	5+++	5+++	0	100.0%		7	7	41.7%	41.7%	
	8001, 12001	6+++	6+++	0	100.0%		15	15	28.6%	28.6%	
	14001, 40001	9+++	8+++	1	88.9%		39	39	17.0%	17.0%	
D03-51213	12001, 18001	6	5	1	83.3%		21	13	19.2%	27.8%	
	6001, 12001	13	5	8	38.5%		57	55	8.1%	8.3%	
	4001, 6001	7	7	0	100.0%		67	67	9.5%	9.5%	
	10001, 18001	6	1	5	16.7%		25	25	3.8%	3.8%	
	16001, 16001	21	7	14	33.3%		32	32	17.9%	17.9%	
	AVERAGE				**89.8%**		**26.3**	**25.7**	**35.8%**	**41.1%**	

(For an explanation of rows indicated by a marker in the "*" column, refer to section 5.3.2)

Table 5.15: Summary of fluid-filled alveoli detection results for all images (Part 2)

		Fluid-Filled Alveoli						
			Within all alveoli					
Slide	Region	Number of Areas	TN	FP	Positive Predictive Value	Specificity	*	Percentage of Alveolar Space considered Fluid-Filled
D03-10538	12001, 20001	30	0	19	61.2%	0.0%		68.8%
	12001, 36001	8+++	1	11	42.1%	8.3%		62.8%
	18001, 10001	1	0	2	33.3%	0.0%		68.1%
	18001, 40001	17	1	11	60.7%	8.3%		72.6%
	30001, 40001	3	0	2	60.0%	0.0%		80.1%
D03-00716	18001, 26001	many	0	17	?	0.0%		57.7%
	22001, 52001	2+++	0	0	100.0%	undef.		71.7%
	14001, 24001	31	0	20	59.2%	0.0%		67.1%
	8001, 22001	0+++	0	24	?	0.0%		76.0%
	34001, 28001	1+++	1	15	?	6.3%		58.0%
D03-45740	8001, 22001	26	0	21	55.3%	0.0%		81.3%
	18001, 42001	many	0	5	?	0.0%		51.6%
	32001, 20001	1	0	34	2.9%	0.0%		65.2%
	14001, 20001	26	0	6	81.3%	0.0%		79.9%
	6001, 20001	3	0	0	100.0%	undef.		65.3%
D03-11282	4001, 24001	49	0	2	96.1%	0.0%		71.2%
	2001, 30001	49+++	0	3	94.1%	0.0%		80.8%
	24001, 46001	0	23	13	0.0%	63.9%		32.5%
	14001, 22001	14	5	62	18.4%	7.5%		51.7%
	26001, 34001	11	63	3	62.5%	95.5%		25.0%
D96-4925	18001, 2001	2	0	21	8.7%	0.0%		64.7%
	24001, 26001	0	0	16	0.0%	0.0%		67.6%
	12001, 32001	4+++	0	5	?	0.0%		55.5%
	2001, 16001	17+++	0	30	36.2%	0.0%		61.6%
	16001, 44001	20+++	0	13	60.6%	0.0%		62.5%
D03-51216	6001, 20001	67+++	0	21	76.1%	0.0%		68.3%
	10001, 30001	13	0	27	32.5%	0.0%		61.5%
	8001, 4001	15	0	23	37.8%	0.0%		56.8%
	8001, 28001	19+++	0	1	95.0%	0.0%		56.1%
	4001, 12001	51+++	0	2	96.2%	0.0%		74.3%
D03-37303	12001, 28001	0+++	7	23	?	23.3%		64.9%
	24001, 16001	3	0	45	6.3%	0.0%		65.4%
	22001, 26001	0+++	0	10	?	0.0%		70.9%
	8001, 42001	3+++	0	0	100.0%	undef.		72.0%
	2001, 42001	1+++	0	0	100.0%	undef.		81.0%
D03-36878	1, 22001	15	0	59	13.2%	0.0%		56.7%
	18001, 14001	34	0	52	39.5%	0.0%		58.4%
	18001, 8001	38	0	74	32.7%	0.0%		65.9%
	22001, 12001	18	0	17	51.4%	0.0%		83.8%
	20001, 6001	17	0	93	15.5%	0.0%		51.4%
N95-4825	20001, 10001	11+++	0	41	21.2%	0.0%		47.3%
	10001, 40001	19+++	12	12	57.1%	50.0%		33.1%
	22001, 24001	5+++	1	6	45.5%	14.3%		53.9%
	8001, 12001	6+++	0	13	31.6%	0.0%		65.9%
	14001, 40001	9+++	7	37	17.8%	15.9%		46.2%
D03-51213	12001, 18001	6	30	3	62.5%	90.9%		19.6%
	6001, 12001	13	11	54	8.5%	16.9%		57.5%
	4001, 6001	7	9	66	9.6%	12.0%		51.9%
	10001, 18001	6	28	17	5.6%	62.2%		18.9%
	16001, 16001	21	70	21	25.0%	76.9%		18.8%
	AVERAGE			**21.4**	**41.9%**	**12.00%**		

5.3.2 Discussion of Results

5.3.2.1 All Alveoli

The detection rate for the "All Alveoli" category is extremely good, with an average sensitivity of 94.8%. Sensitivity scores within this category were above 85% for 48 of the 50 images.

The two image regions with sensitivity below 85% were both from the same slide: D03-51213. The image of this slide with upper left corner coordinates [6001, 12001] had a sensitivity value of 75.6%. In this case, the image had a total of 78 alveoli marked in the manual analysis, 65 of which were air-filled, and 13 of which were fluid-filled. The automated analysis detected 53 of the 65 air-filled alveoli, but failed to find the remaining 12. All 12 of these missed alveoli are extremely small, and are below the minimum alveoli size cutoff used by the algorithms. They were not marked by the automated analysis for this reason. Of the fluid-filled alveoli, five of the 13 were found, and eight missed. These eight alveoli all appear extremely dark pink in colour. The algorithms incorrectly classified these alveoli as part of the category containing alveolar walls, fibrin and necrosis, during the RGB clustering stage.

The second image region with a lower sensitivity value has upper left corner co-ordinates [10001, 18001]. The sensitivity value for this region was 82.4%. In this case, the image contained a total of 51 alveoli of which 45 were considered air-filled. The automated analysis successfully detected 41 of the 45 air-filled alveoli. One of the four alveoli that were missed lies on the border of the image region, at the corner. The software incorrectly considered this alveolus to be part of the background. Misclassifying alveoli which lie at the corner of image regions in this way is a known flaw in the thesis software (see section 4.1 for details). The three remaining air-filled alveoli which were missed are all extremely small, and are below the minimum alveolus size cutoff of the programs. Results for the fluid-filled alveoli of this image are not as good. Here only one of six fluid-filled alveoli was detected. The remaining five alveoli are all extremely dark pink in colour, so, as with the previous image, they were incorrectly classified during the RGB clustering stage.

The number of false positives in the detection of alveoli is generally very low, with an average of 6.7 false positives per image region. (When false positives within areas labeled "not to be analyzed" are excluded, the adjusted false positive value drops to 5.4 per image region). Eleven of the 50 images had more than 10 false positives each, however. The main reason for

these high false positive scores is once again due to the minimum alveolus size. The thesis software has a set minimum alveolus size of 1000 pixels. Any detected alveoli below this size are ignored by the software. The 1000 pixel minimum size was selected to correspond with the smallest alveoli that would be marked in the manual analysis. However, judging alveoli sizes is much more difficult in the manual analysis, and it appears that very small alveoli were not treated consistently on all manually analyzed image regions. For some, such as the two images of D03-51213 which had lower sensitivity scores, very small alveoli below the minimum size cutoff were still marked in the manual analysis. For other images, very small alveoli which were slightly above the 1000 pixel cutoff size were not marked in the manual analysis, resulting in false positives when these small alveoli were marked by the automated analysis. For the ten images with high false positive counts for alveolar detection, this appears to be the main reason.

5.3.2.2 Fluid-Filled Alveoli

False positive scores for fluid-filled alveoli were much higher than those of all alveoli for the same image regions. False positives for fluid-filled alveoli include both other tissue areas (such as fibrin) which were incorrectly identified as fluid-filled alveoli, as well as air-filled alveoli which were correctly detected but incorrectly labeled as fluid-filled by the programs. The mislabeling of the contents of correctly identified alveoli is the largest contributor to the high false positive rates. The average false positive rate for fluid-filled alveoli within the full set of objects in the image is 26.3. When all objects other than alveoli are excluded, and the false positives for fluid-filled alveoli counted within the set of alveoli only, this value drops only to an average value of 21.4.

Specificity was measured for fluid-filled alveoli within the set of alveoli. For 31 of the 50 analyzed image regions, specificity was 0%, indicating that all air-filled alveoli in the image had been incorrectly labeled as containing fluid. There are a number of reasons for the high rate of misidentified alveolar contents. In the automated analysis, any alveolus which had more than 33% of its pixels classified as pink was considered fluid-filled. Although this same standard was applied to the manual analysis also, it is much more difficult to judge the exact percentage of pink area visually. Results show that many more alveoli were considered fluid-filled by the automated analysis than by the manual analysis. It is possible, and very likely, that some alveoli

which had a percentage of pink pixels of slightly more than 33% were classified as fluid-filled by the automated analysis, but as air-filled by the manual analysis.

Measuring the percentage of pink pixels is also not necessarily the best way to assess the alveolar contents, if the goal is to match manual assessment results. In some cases, an alveolus will contain regions of air and regions of fluid. In this case, manually assessing the percentage of fluid is relatively easy. However, in other cases, pink pixels may be evenly dispersed in the alveolar region. This would correspond to a situation where the alveolus does contain some fluid, but not enough to cause a darker staining result. Where a third of the pixels are pink, but these pixels are evenly distributed among two-thirds white pixels, the alveolus may appear to the eye to be white, corresponding to an air-filled area. The programs assess the contents numerically, and would still assign a fluid-filled label to this alveolus.

An additional cause of high fluid-filled alveoli numbers in the automated analysis stems from issues with determining alveolar boundaries. In some cases, the automated analysis judged several alveolar areas to be connected, and identified the larger region as a single alveolus, whereas the manual analysis showed the same region as consisting of a number of separate alveoli. If only one or a few of these connected alveolar areas is pink, the percentage of pink pixels in the entire area may still be above the 33% threshold, and the automated analysis would label the entire group as fluid-filled. In the manual analysis, only those individual alveoli in the group which were pink would be labeled as fluid-filled, and the others as air-filled. This would result in several false positives for fluid-filled alveoli for the automated analysis. An example of this situation is shown below (Figure 5.37). This example is taken from region [12001, 20001] of slide D03-10538.

In the example shown in Figure 5.37, the central area of this image was counted as a single fluid-filled alveolus in the automated analysis. This same area was marked as five alveoli by the manual analysis: two fluid-filled and three air-filled. Consequently, three false positives for fluid-filled alveoli were scored for this one area.

Although all alveoli in these 31 images were labeled as fluid-filled by the automated analysis, it is interesting to note that the percentage of alveolar area which is considered fluid-filled for these cases ranges from a low of only 47.3% to a high of 83.8%. The average percentage of fluid-filled alveolar space for these images is only 65.9%. This low percentage shows that although alveoli are considered fluid-filled overall, many likely contain significant

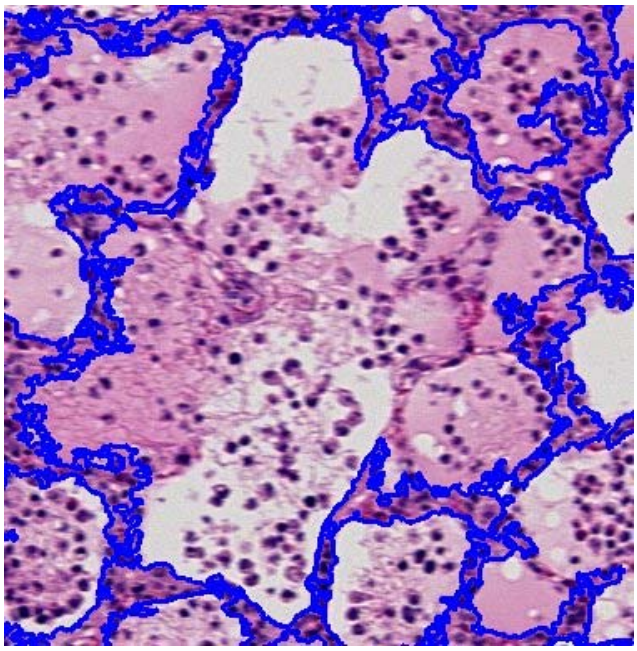


Figure 5.37: Example of an area marked as a single alveolus in the automated analysis, but as multiple alveoli in the manual analysis

air-filled areas also, which may have been marked as separate air-filled alveoli in the manual assessment.

Sensitivity scores for fluid-filled alveoli are generally very good, with values above 80% for 36 of the 50 image regions. Seven of the other images have no sensitivity values recorded. Two of these images had no specific alveoli labeled in the manual analysis but had all background tissue identified as fluid. The automated analysis identified numerous fluid-filled alveoli in both cases, but since there is no known number of actual regions which should be found, the sensitivity values for these two cases are excluded from the average and listed merely as "?". Five additional images have sensitivity listed as "undef.". In these cases, the images either had no alveoli marked and none found by the automated analysis, or there were no specific fluid-filled alveoli marked although there were specific air-filled alveoli identified in the manual analysis. In either case, the number of true positives and false negatives for these cases were both zero, resulting in an undefined sensitivity score. For the remaining seven images with sensitivities below 80%, the sensitivity values were very low, ranging from 60% to 16.7%. Two of these low sensitivity images were the same ones which had relatively low sensitivity scores for all alveoli also: image regions with upper left corners [6001, 12001] and [10001, 18001] of slide D03-51213. As explained previously, low detection rates for fluid-filled alveoli within these two images was due to the very dark pink colour of these alveoli, which resulted in the

alveoli being incorrectly classified by the automated analysis as alveolar walls, fibrin or necrosis. This same problem appears to be the issue with a third image region from the same slide. The area of slide D03-51213 with upper left corner [16001, 16001] has 21 fluid-filled alveoli of which only seven were detected, yielding a sensitivity score of 33.3%. The 14 fluid-filled alveoli which were missed were also an extremely dark pink colour. The fourth image with a low sensitivity score was the region at co-ordinates [26001, 34001] of slide D03-11282. In this case, there were 11 fluid-filled alveoli. All 11 were found, but 6 of these were incorrectly labeled as air-filled, resulting in a sensitivity score for fluid-filled alveoli of only 45.5%. These six alveoli are all very light in colour, and so likely have slightly less than the 33% pink pixels required by the automated analysis to be considered fluid-filled. The final image with a low sensitivity score was region [1, 22001] of slide D03-36878, with a sensitivity value of 60%. In this case, there were 15 fluid-filled alveoli, of which 6 were not found. All six of these alveoli are highly textured, containing numerous visible nuclei. In this case, these areas contained a high number of black pixels in the RGB classified image, and were not isolated as alveoli. The remaining two images with low sensitivity scores are regions [22001, 52001] of slide D03-00716 and [12001, 32001] of slide D96-4925. In both cases, a few specific fluid-filled alveoli were labeled in the manual analysis, along with all background tissue identified as fluid. For image [22001, 52001] of slide D03-00716, two specific fluid-filled alveoli were marked but only one of these was detected by the automated analysis. Consequently, sensitivity based on specific alveoli is only 50% for this image. This figure is misleading however, since many more fluid-filled alveoli were identified in the automated analysis, all of which correspond to the nonspecifically marked background fluid area identified in the manual assessment. The same situation occurs for image [12001, 32001] of slide D96-4925 in which only two of the four specifically marked fluid-filled alveoli were detected by the automated analysis, along with many other fluid-filled alveoli identified, all of which are correct since they overlap with the non-specifically labeled background fluid area. A sensitivity of only 50% is also recorded for this image, based on the countable alveoli.

5.4 Results of Inflammatory Cells Detection

5.4.1 Measurement of Results

In general, scoring detection results for inflammatory cells was more difficult than assessing those of alveoli detection, since regions of inflammatory cells do not necessarily have distinct boundaries. As was the case with fluid, in a number of images the full tissue area was identified as containing infiltrations of these cells in the manual assessment, with or without any specific areas of inflammatory cells identified. The automated analysis, on the other hand, did outline specific areas of inflammatory cells in all cases where any were found. Fourteen of the 50 image regions had areas of inflammatory cells labeled this way in the manual assessment. For those image regions where some specific areas of inflammatory cells were identified, along with all unlabeled areas being identified as containing these cells, the true positive count was limited to the number of specific areas marked in the manual analysis, even though this number is much lower than the actual value should be. In these cases, the number listed in the "Number of Areas" column of the results chart (Table 5.16) includes a "+++" symbol, indicating that many more areas of inflammatory cells existed. In cases where all tissue area was identified as containing inflammatory cells with no specific areas of infiltration of these cells identified, the "Number of Areas" amount is merely listed as "many". The exact number of true positives and false negatives for these cases is not known. The true positives could be set to the number of areas of inflammatory cells identified in the automated analysis, in which case the false negative score would have to be zero. The resulting 100% sensitivity score would be misleading, however. Instead, no sensitivity value based on countable areas of inflammatory cells is included for these images, and "n/a" is listed in the results chart. This situation occurs for two image regions: area with upper left co-ordinates [24001, 26001] of slide D96-4925, and area with co-ordinates [22001, 24001] of slide N95-4825.

The most difficult situation for calculating sensitivity based on countable regions occurs where all tissue area is identified as containing inflammatory cells, some specific areas of infiltration of these cells are also marked, and some of these specific areas are not detected by the automated analysis. In these cases, the number of false negatives is known, but the number of true positives is not. Setting the true positive number will affect the sensitivity score. It does not seem correct to set this number to the number of regions identified in the automated analysis. In these cases, the false negative number is identified, but the true positive count and the sensitivity

are merely listed as "?" in the results chart. This situation occurs for two slide image regions: areas with upper left corners [8001, 22001] and [34001, 28001] of slide D03-00716. The sensitivity scores for these two image regions are thus not included in the average sensitivity for detection of areas of inflammatory cells. Since it is reasonable to expect a very high sensitivity value (the number of true positive areas would be much higher than the low number of false negative for these image regions), the average sensitivity is thus likely to be slightly lower than it should be.

Due to the difficulties in assessing results for image regions where all tissue area is non-specifically identified as containing inflammatory cells, an additional measure is used to evaluate these outcomes. The automated assessment produces quantitative data, some of which describes the percentage of the overall image area which is considered infiltrated by inflammatory cells. Comparing the results based on areas, not specific regions, seems to be a more useful technique in these cases. The challenge with this approach, however, is that the manual assessment does not identify the amount of area occupied by inflammatory cells. Where the manual assessment identified all tissue area in an image as containing these cells, except for individually identified alveoli, areas not to be analyzed, regions of fibrin and necrosis etc. (i.e. all "background" tissue contains inflammatory cells), it would be necessary to calculate the total tissue area minus the areas of these specific features in order to compute the area occupied by inflammatory cells. It is certainly possible to do this, but it is also somewhat challenging and time-consuming due to the irregular shapes of all of these areas. These calculations for the manual assessment have not been done at this stage. As a simplification, areas are compared by assuming that in the manual assessment inflammatory cells occupy 100% of the tissue area in images where all "background" tissue was labeled as containing inflammatory cells. That is, all specifically labeled areas are not subtracted from this amount. The specific area infiltrated by inflammatory cells identified by the automated analysis is then compared to this 100% figure. Obviously, this will result in lower sensitivity scores than are accurate. The results table lists the area occupied by inflammatory cells, as assessed by the automated analysis, for each image region. Sensitivity scores based on a comparison of areas is only included for those image regions which had non-specific background areas labeled as containing inflammatory cells. In images where all identified inflammatory cells are located within specifically marked regions, results can be evaluated based on detection of these regions, and the entry for sensitivity based on area is listed as "n/a" for these cases.

Finally, one additional column has been added to the results chart for inflammatory cells. The “Sensitivity calculated using most appropriate method” column contains the sensitivity score based on countable regions for all images where this value could be computed. For those regions where calculating sensitivity based on individual regions is not possible, this column contains the sensitivity score based on area.

Table 5.16: Summary of inflammatory cells detection results for all images

		Areas of Inflammatory Cells												
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*	Percentage of Tissue Area	Sensitivity based on Area	Sensitivity calculated using most appropriate method
D03-10538	12001, 20001	12	12	0	100.0%		25	25	32.4%	32.4%		36.21	n/a	100.0%
	12001, 36001	7+++	7	0	100.0%		0	0	100.0%	100.0%		45.21	45.2%	45.2%
	18001, 10001	5	5	0	100.0%		6+	0	45.5%	100.0%		38.41	n/a	100.0%
	18001, 40001	7+++	7	0	100.0%		1	0	87.5%	100.0%		45.96	46.0%	46.0%
	30001, 40001	5	4	1	80.0%		8+	5+	33.3%	44.4%		33.53	n/a	80.0%
D03-00716	18001, 26001	23	23	0	100.0%		1	0	95.8%	100.0%		90.84	n/a	100.0%
	22001, 52001	14	14	0	100.0%		2	0	87.5%	100.0%		63.99	n/a	100.0%
	14001, 24001	6	6	0	100.0%		4+	1+	60.0%	85.7%		86.00	n/a	100.0%
	8001, 22001	4+++	0?	4	?	??	1	0	?	?	??	2.75	2.8%	2.8%
	34001, 28001	1+++	0?	1	?	??	0	0	?	?	??	10.46	10.5%	10.5%
D03-45740	8001, 22001	1	1	0	100.0%		8+	8+	11.1%	11.1%		27.76	n/a	100.0%
	18001, 42001	0	0	0	undef.		10+	1+	0.0%	0.0%		32.72	n/a	undef.
	32001, 20001	11	1	10	9.1%		10+	10+	9.1%	9.1%		22.93	n/a	9.1%
	14001, 20001	4	4	0	100.0%		9+	9+	30.8%	30.8%		33.76	n/a	100.0%
	6001, 20001	2	2	0	100.0%		6+	2+	25.0%	50.0%		45.02	n/a	100.0%
D03-11282	4001, 24001	5+++	5	0	100.0%		8+	5+	38.5%	50.0%		41.51	41.5%	41.5%
	2001, 30001	0	0	0	undef.		20+	18+	0.0%	0.0%		33.85	n/a	undef.
	24001, 46001	0	0	0	undef.		16	15	0.0%	0.0%		7.18	n/a	undef.
	14001, 22001	3	3	0	100.0%		6+	6+	33.3%	33.3%		41.51	n/a	100.0%
	26001, 34001	0	0	0	undef.		25+	25+	0.0%	0.0%		18.21	n/a	undef.
D96-4925	18001, 2001	28	26	2	92.9%		2	0	92.9%	100.0%		71.40	n/a	92.9%
	24001, 26001	many	many	?	n/a	??	12	5	?	?	??	40.21	40.2%	40.2%
	12001, 32001	14	14	0	100.0%		6	0	70.0%	100.0%		62.49	n/a	100.0%
	2001, 16001	27	0	27	0.0%	**	1	1	0.0%	0.0%		0.36	n/a	0.0%
	16001, 44001	52+++	49+++	3	94.2%		5	0	90.7%	100.0%		51.36	51.4%	51.4%
D03-51216	6001, 20001	9+++	9	0	100.0%		13	9	40.9%	50.0%		43.43	43.4%	43.4%
	10001, 30001	2	2	0	100.0%		14	12	12.5%	14.3%		28.27	n/a	100.0%
	8001, 4001	1	1	0	100.0%		8+	8+	11.1%	11.1%		56.98	n/a	100.0%
	8001, 28001	5+++	5	0	100.0%		9	6	35.7%	45.5%		36.88	36.9%	36.9%
	4001, 12001	5+++	5	0	100.0%		3	3	62.5%	62.5%		19.26	19.3%	19.3%
D03-37303	12001, 28001	1	0	1	0.0%		8	5	0.0%	0.0%		4.32	n/a	0.0%
	24001, 16001	2	2	0	100.0%		2	2	50.0%	50.0%		69.23	n/a	100.0%
	22001, 26001	3	3	0	100.0%		6+	6+	33.3%	33.3%		70.01	n/a	100.0%
	8001, 42001	1	0	1	0.0%		2	0	0.0%	undef.		0.40	n/a	0.0%
	2001, 42001	0	0	0	undef.		10+	10+	0.0%	0.0%		12.49	n/a	undef.
D03-36878	1, 22001	7	6	1	85.7%		13+	10+	31.6%	37.5%		46.82	n/a	85.7%
	18001, 14001	0	0	0	undef.		14+	14+	0.0%	0.0%		29.23	n/a	undef.
	18001, 8001	0	0	0	undef.		12+	12+	0.0%	0.0%		22.65	n/a	undef.
	22001, 12001	1	1	0	100.0%		12+	9+	7.7%	10.0%		23.14	n/a	100.0%
	20001, 6001	0	0	0	undef.		9+	9+	0.0%	0.0%		17.66	n/a	undef.
N95-4825	20001, 10001	1+++	1	0	100.0%		0	0	100.0%	100.0%		68.13	91.7%	91.7%
	10001, 40001	1+++	1	0	100.0%		0	0	100.0%	100.0%		37.54	54.2%	54.2%
	22001, 24001	many	many	?	n/a	??	0	0	100.0%	100.0%		70.32	82.8%	82.8%
	8001, 12001	3	3	0	100.0%		11+	9+	21.4%	25.0%		63.29	n/a	100.0%
	14001, 40001	3+++	3	0	100.0%		4+	0	42.9%	100.0%		38.06	50.9%	50.9%
D03-51213	12001, 18001	0	0	0	undef.		8+	6+	0.0%	0.0%		37.65	n/a	undef.
	6001, 12001	0	0	0	undef.		9+	7+	0.0%	0.0%		40.29	n/a	undef.
	4001, 6001	0	0	0	undef.		4+	3+	0.0%	0.0%		85.32	n/a	undef.
	10001, 18001	0	0	0	undef.		8+	8+	0.0%	0.0%		39.58	n/a	undef.
	16001, 16001	0	0	0	undef.		7+	7+	0.0%	0.0%		53.14	n/a	undef.
	AVERAGE				**86.7%**		**7.6**	**5.8**	**33.9%**	**41.0%**			**44.0%**	**67.1%**

(For an explanation of rows indicated by a marker in the "*" column, refer to section 5.4.2)

5.4.2 Discussion of Results

Average sensitivity for inflammatory cells detection is 86.7%, based on 33 image regions, when using the countable regions method. Thirteen image regions are excluded because they had no areas of inflammatory cells identified in the manual analysis, resulting in both TP and FN scores of 0, and sensitivity scores that are undefined. (The entry in the sensitivity column for these cases is shown as "undef."). An additional four image regions are excluded since calculating sensitivity based on countable regions is not possible for these images. These four cases are examined in detail below.

There were two image regions for which the number of areas of inflammatory cells is listed merely as "many" in the results chart. For both region [24001, 26001] of slide D96-4925 and region [22001, 24001] of slide N95-4825, all unmarked tissue areas were labeled as containing inflammatory cells in the manual analysis. In the automated analysis for both image regions, numerous areas of inflammatory cells were identified, covering a large amount of the tissue area. No sensitivity count based on detection of regions of inflammatory cells was included for either of these images since the number of true positives and false negatives is not known, as explained previously. Both regions [8001, 22001] and [34001, 28001] of slide D03-00716 posed an interesting problem for calculating sensitivity. In both cases, a few specific areas of inflammatory cells were identified in the manual analysis along with all tissue being labeled as containing these cells. In both cases, the automated analysis did identify some areas of inflammatory cells in the images, but did not find the specific areas labeled in the manual analysis. Based on a count of regions, the true positive score would be 0, and sensitivity would be 0%. However, all areas labeled as inflammatory cells in these images by the automated analysis are correct, since they overlap with the background area which was also labeled as containing inflammatory cells. It seems neither correct to base a true positive count on the number of these regions marked by the automated analysis, nor to count TP as zero and ignore these correctly labeled regions. For both of these images, TP is listed as '?' and sensitivity is also listed as '?' in the results chart.

Sensitivity scores based on countable regions are very high for the majority of image regions which contain inflammatory cells. Twenty-nine of these 37 images have sensitivity scores of 80% or greater. Two additional images are the cases where the number of regions of inflammatory cells is merely listed as "many", and no specific sensitivity value is calculated, as

discussed above. For both of these images, numerous areas of inflammatory cells were also detected by the automated analysis, so sensitivity is not an issue. In contrast, three images which contain areas of inflammatory cells had none of these specific areas identified by the automated analysis. This failure to detect inflammatory cells is most significant for image region [2001, 16001] of slide D96-4925. In this case, 27 areas of these cells were marked in the manual analysis, and none of them were found by the automated analysis. The automated analysis did identify one very small area of inflammatory cells in this image, occupying a total of only 0.36% of the image area. This region of inflammatory cells did not overlap with any of the 27 areas identified in the manual analysis, and was counted as a false positive. If diagnostic conclusions were drawn from the automated analysis results alone, one would assume that infiltration of neutrophils and/or macrophages is not a significant factor in the disease state of this sample. Visually, however, this image has large areas of densely packed nuclei, which is exactly the feature the inflammatory cells detection algorithm should have found. More analysis is required to determine why the algorithm did not detect areas of inflammatory cells in this case. For the other two images with 0% sensitivity scores based on countable regions of inflammatory cells, only a single area was marked in the manual analysis of each image and was missed by the automated analysis. In both cases, the detection algorithm did identify some areas of inflammatory cells in the images, in regions which did not overlap with those indicated by the manual analysis. The amount of area of inflammatory cells marked by the automated analysis for each of these images is low, which is consistent with the fact that only a single region of these cells was marked in each by the manual analysis. For image [12001, 28001] of slide D03-37303, a total of eight regions of inflammatory cells were identified by the automated analysis, totaling 4.32% of the image area. For image [8001, 42001] of slide D03-37303, two areas of inflammatory cells were found, covering 0.40% of the image area. The remaining low sensitivity score was for region [32001, 20001] of slide D03-45740. For this image, only one of 11 areas of inflammatory cells was detected, resulting in a sensitivity of only 9.1%. There is no obvious reason why these areas of inflammatory cells were missed. More than 10 other areas of these cells were identified in this image by the automated analysis, however, covering a total of 22.93% of the image area. Overall, although sensitivity values are low for four images, there is only one case in any of the 37 in which areas of inflammatory cells exist where the automated analysis results indicate that infiltration of neutrophils and/or macrophages is not a factor.

Sensitivity based on area is included for each of the 14 image regions in which all tissue was nonspecifically labeled as containing inflammatory cells in the manual assessment, whether or not individual areas were also marked. In most of these cases, since no area figure is easily obtainable for the manual analysis, it is assumed that 100% of the tissue area was identified as containing inflammatory cells in this analysis. This 100% figure is compared with the percentage of image area identified as containing inflammatory cells in the automated analysis. The exception is for the four image regions of slide N95-4825. Several large air-filled alveoli are present in each of the image regions of this slide. The air-filled alveoli are excluded from the generic tissue area which was labeled as inflammatory cells in the manual analysis. There is also unlikely to be any overlap of air-filled regions of alveoli with areas of inflammatory cells in the automated analysis. The programs which locate and label alveoli are separate from those which identify and label areas of inflammatory cells, and so there is no guarantee that the same image area will not be labeled by both algorithms. However, any portion of alveolar area which could possibly be labeled as inflammatory cells by the detection algorithm was classified as non-white pixels by the alveoli detection algorithm, and as a result counted as fluid by this algorithm. There could potentially be some amount of overlap of the areas identified as fluid by the alveoli detection program and those identified as inflammatory cells by the inflammation detection algorithm, but there should be absolutely no overlap of the pixels identified as air with those identified as inflammatory cells. For the four image regions of slide N95-4825, the amount of alveolar area classified as air-filled was subtracted from the overall tissue area, and then the percentage of tissue area identified as inflammatory cells was compared to the remainder to calculate the sensitivity based on area. For example, for image region [20001, 10001] of slide N95-4825, alveolar space occupied 1,957,561 of the 4,000,000 pixels of this image. Of this area, 47.3% was judged to contain fluid, and the remaining 52.7% to contain air. Consequently, air-filled alveolar space occupied 25.8% of the total image area. Assuming that the manual analysis identified only the remaining 74.2% of the tissue as containing inflammatory cells, the 68.1% of tissue area identified as inflammatory cells by the automated analysis is compared to this figure, resulting in a sensitivity based on area of 91.7%. For image regions other than those for slide N95-4825, the amount of alveolar space which was considered air-filled was negligible, and it could not be safely assumed that this area would have been excluded from the non-specific tissue labeled as containing inflammatory cells in the manual assessment. As a result, the percentage

of tissue area that was identified as containing inflammatory cells in the automated assessment was compared to 100% to calculate the sensitivity based on area for these images. These sensitivity values are lower than is accurate, since if the exact amount of tissue area identified as inflammatory cells in the manual analysis was calculated, this value would be less than the 100% figure which was used.

Sensitivity scores based on area range from a low of only 2.8% to a high of 91.7%. The average sensitivity based on area for the 14 image regions for which this figure is included is 44.0%. The two images with particularly low sensitivity based on area scores are both from slide D03-00716. For region [8001, 22001], four small areas of inflammatory cells were found by the automated analysis, covering only 2.8% of the tissue area. All tissue in this image which was not specifically identified was labeled as a combination of fluid, fibrin and inflammatory cells by the manual analysis. Based on this non-specific labeling, it is reasonable to assume that comparing the amount of area of inflammation with 100% of tissue area is very inaccurate, and yields a much lower sensitivity value than should be recorded. For region [34001, 28001] of the same slide, 10.5% of the tissue area was identified as inflammatory cells by the automated analysis. In the manual analysis, the background tissue was labeled as a combination of fluid and "limited amounts of inflammatory cells". Obviously assuming that 100% of the tissue area was identified as containing inflammatory cells in the manual analysis leads to a much lower sensitivity score than is accurate in this case.

When examining specificity of the inflammatory cells detection algorithm, counting the number of false positives is arguably of little value. If inflammatory cells are identified in areas where none exist, this is relevant. However, whether three small regions of these cells are identified within this area, or whether the area is marked as a single region of inflammatory cells, seems of little consequence. A more meaningful measurement would be the amount of area incorrectly identified as containing inflammatory cells. Since the agreed upon technique for this project was to compare and count specific regions, an attempt has been made to produce these false positive values. In cases where the automated analysis marked numerous very small areas of inflammatory cells along with some larger areas, only the more significantly sized areas were counted. The number of false positives listed includes a "+" marker in these cases.

Useful information can be gathered by making some observations about the occurrence of false positives, even if exact numbers are not examined. For example, it is important to note

cases where any areas of inflammatory cells were identified by the automated analysis where none were found in the manual analysis. This was the case for 13 image regions. The amount of area identified as containing inflammatory cells was relatively small, however, for most of these regions. For seven image regions containing no inflammatory cells in slides other than D03-51213, the areas varied from a low of 7.2% of the tissue area up to 33.9% of the tissue area. Slide D03-51213 posed a problem for the automated analysis of inflammatory cells, however. All five image regions of this slide were identified in the manual analysis as not containing inflammatory cells, but regions of these cells were indicated in all of these images by the automated analysis, covering from 37.7% to 85.3% of the tissue area. The problem in this case is that the detection algorithm is misinterpreting atelectasis as an infiltration of neutrophils and/or macrophages. In all five images of slide D03-51213, alveoli have collapsed bringing the epithelial cells of the alveolar walls close together. The closely spaced nuclei of these epithelial cells are detected and incorrectly identified as clumps of inflammatory cells by the detection algorithm.

Although the results discussed above all indicate issues with the specificity of inflammatory cells detection, there were also numerous results where the algorithm was extremely specific. The only images which had zero false positives for areas of inflammatory cells were those in which all background area was labeled as containing these cells in the manual analysis. In these cases, any area identified as inflammatory cells was correct, provided that it did not overlap with areas which were identified as other feature types. Only four image regions had a false positive count of zero. When areas identified as inflammatory cells within regions marked “not to be analyzed” were excluded, however, a total of 15 image regions had Adjusted False Positive counts of zero. Overall, nine image regions had a Positive Predictive Value of greater than 80%, and 13 image regions had an Adjusted Positive Predictive Value of greater than 80%.

When all results are included however, average Positive Predictive Value for detection of areas of inflammatory cells is only 33.9%, with this figure rising slightly to 41.0% for average Adjusted Positive Predictive Value.

5.5 Results of Fibrin Detection

5.5.1 Measurement of Results

As was the case with inflammatory cells also, for some slides the entire tissue area was labeled as fibrin in the manual analysis, with or without individual areas of fibrin marked. Consistent with the handling of these cases for the inflammatory cells results, where the entire tissue area was identified as fibrin, but some specific areas of fibrin were also identified, the value in the "Number of Areas" column (Table 5.17) indicates the number of specific areas of fibrin. A "+++" symbol is included to indicate that there are actually many more than the specific number of areas of fibrin listed. This was the case for four images: regions with upper left co-ordinates [8001, 22001], [32001, 20001], and [14001, 20001] of slide D03-45740, and region [8001, 42001] of slide D03-37303. For images where the entire background tissue is identified as fibrin with no specific areas of fibrin marked, the value in the "Number of Areas" column is merely listed as "many". This was the case for three images: region [8001, 22001] of slide D03-00716, region [6001, 20001] of slide D03-45740, and region [2001, 42001] of slide D03-37303. Unlike inflammatory cells, however, the number of areas of fibrin identified by the automated analysis is easily countable. For the True Positive tally, any specific areas of fibrin identified by the manual analysis which were detected by the automated analysis are counted, and any areas of fibrin identified by the automated analysis which correspond to the background fibrin area are also included. Scoring sensitivity based only on these countable regions is inadequate. The column "Percentage of Tissue Area" is included in the Results chart to show the amount of tissue area identified by the automated analysis as fibrin. Sensitivity can also be measured based on this area figure, but this technique also has shortcomings, as discussed in the next section.

Table 5.17: Summary of fibrin detection results for all images

		Fibrin										
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*	Percentage of Tissue Area
D03-10538	12001, 20001	25	3	22	12.0%		0	0	100.0%	100.0%		8.3
	12001, 36001	23	11	12	47.8%		1	1	91.7%	91.7%		26.9
	18001, 10001	3	1	2	33.3%		9	5	10.0%	16.7%		17.5
	18001, 40001	22	12	10	54.5%		2	2	85.7%	85.7%		11.5
	30001, 40001	23	6	17	26.1%		1	0	85.7%	100.0%		7.1
D03-00716	18001, 26001	3	2	1	66.7%		2	2	50.0%	50.0%		2.3
	22001, 52001	0	0	0	undef.		6	6	0.0%	0.0%		3.1
	14001, 24001	4	2	2	50.0%		1	1	66.7%	66.7%		5.2
	8001, 22001	many	2	0+++	100.0%	??	4	3	33.3%	40.0%		6.4
	34001, 28001	0	0	0	undef.		1	1	0.0%	0.0%		0.5
D03-45740	8001, 22001	2+++	10	0	100.0%		4	2	71.4%	83.3%		11.2
	18001, 42001	0	0	0	undef.		10	2	0.0%	0.0%		28.1
	32001, 20001	5+++	17	3	85.0%		0	0	100.0%	100.0%		12.8
	14001, 20001	2+++	11	0	100.0%		2	2	84.6%	84.6%		22.8
	6001, 20001	many	3	0	100.0%	??	11	1	21.4%	75.0%		21.5
D03-11282	4001, 24001	15	9	6	60.0%		6	1	60.0%	90.0%		13.6
	2001, 30001	17	5	12	29.4%		8	7	38.5%	41.7%		30.0
	24001, 46001	0	0	0	undef.		4	3	0.0%	0.0%		13.7
	14001, 22001	2	0	2	0.0%		0	0	undef.	undef.		0.0
	26001, 34001	0	0	0	undef.		4	4	0.0%	0.0%		2.8
D96-4925	18001, 2001	3	0	3	0.0%		0	0	undef.	undef.		0.0
	24001, 26001	1	1	0	100.0%		4	1	20.0%	50.0%		8.2
	12001, 32001	16	0	16	0.0%		0	0	undef.	undef.		0.0
	2001, 16001	2	0	2	0.0%		2	2	0.0%	0.0%		0.4
	16001, 44001	0	0	0	undef.		1	1	0.0%	0.0%		0.3
D03-51216	6001, 20001	14	11	3	78.6%		3	1	78.6%	91.7%		11.5
	10001, 30001	23	7	16	30.4%		1	0	87.5%	100.0%		14.8
	8001, 4001	15	3	12	20.0%		4	4	42.9%	42.9%		17.3
	8001, 28001	23	13	10	56.5%		3	3	81.3%	81.3%		22.7
	4001, 12001	23	10	13	43.5%		4	4	71.4%	71.4%		7.9
D03-37303	12001, 28001	15	1	14	6.7%		7	5	12.5%	16.7%		3.5
	24001, 16001	5	1	4	20.0%		0	0	100.0%	100.0%		7.8
	22001, 26001	8	0	8	0.0%		1	1	0.0%	0.0%		0.4
	8001, 42001	1+++	1	1	50.0%	??	0	0	100.0%	100.0%		1.2
	2001, 42001	many	10	0	100.0%	??	0	0	100.0%	100.0%		9.2
D03-36878	1, 22001	11	1	10	9.1%		1	1	50.0%	50.0%		0.9
	18001, 14001	3	1	2	33.3%		3	3	25.0%	25.0%		2.3
	18001, 8001	0	0	0	undef.		0	0	undef.	undef.		0.2
	22001, 12001	7	6	1	85.7%		7	7	46.2%	46.2%		12.7
	20001, 6001	0	0	0	undef.		4	4	0.0%	0.0%		1.5
N95-4825	20001, 10001	2	0	2	0.0%		0	0	undef.	undef.		0.0
	10001, 40001	0	0	0	undef.		2	1	0.0%	0.0%		1.6
	22001, 24001	15	0	15	0.0%		0	0	undef.	undef.		0.0
	8001, 12001	12	4	8	33.3%		2	2	66.7%	66.7%		2.4
	14001, 40001	0	0	0	undef.		7	1	0.0%	0.0%		3.5
D03-51213	12001, 18001	0	0	0	undef.		3	1	0.0%	0.0%		4.2
	6001, 12001	0	0	0	undef.		6	0	0.0%	undef.		7.9
	4001, 6001	0	0	0	undef.		3	2	0.0%	0.0%		0.9
	10001, 18001	1	1	0	100.0%		3	2	25.0%	33.3%		2.6
	16001, 16001	0	0	0	undef.		6	6	0.0%	0.0%		2.8
	AVERAGE				**45.3%**		**3.1**	**2**	**41.0%**	**45.5%**		

(For an explanation of rows indicated by a marker in the "*" column, refer to section 5.5.2)

5.5.1 Discussion of Results

Average sensitivity for fibrin detection was 45.3%, based on 36 image regions. No fibrin was present in the remaining 14 image regions, according to the expert manual analysis.

Three image regions had all background tissue identified as fibrin in the manual analysis, with no specific areas of fibrin marked. For all three of these cases (region [8001, 22001] of slide D03-00716, region [6001, 20001] of slide D03-45740, and region [2001, 42001] of slide D03-37303), the False Negative score based on the countable areas technique is zero, and consequently the sensitivity score is 100% regardless of the number of actual areas of fibrin identified by the automated analysis. As was discussed in the inflammatory cells results, this is not an adequate method of assessing sensitivity, and a better technique would be to compare the amount of tissue area which was labeled as fibrin by both analyses. The "Percentage of Tissue Area" column identifies the amount of tissue area assessed as fibrin by the automated analysis. For image area with upper left corner co-ordinates [8001, 22001] of slide D03-00716, a total of only 6.4% of the tissue area was identified as fibrin by the automated analysis. If, as a simplification, it is assumed that the percentage of tissue area identified as fibrin by the manual analysis is 100%, this results in a sensitivity value of only 6.4%. The actual sensitivity score would be higher, since not all tissue area should be included in the amount judged to be fibrin by the manual analysis. In fact, for this particular image region, all tissue areas which were not specifically marked were labeled as a combination of fluid, fibrin, and inflammatory cells. Given this unspecific marking, it is difficult to assess an accurate percentage of tissue area to use in the comparison of automated and manual results based on areas. Neither the sensitivity value obtained by the countable regions technique, nor that obtained by a comparison of areas is accurate, and there is a huge discrepancy between the two values. The same is true for the other two images. For image region [6001, 20001] of slide D03-45740, the total tissue area identified as fibrin by the automated analysis was only 21.5%. If it is assumed that 100% of the tissue area is fibrin according to the manual analysis, the sensitivity score based on areas is only 21.5%. For image region [2001, 42001] of slide D03-37303, a total of 10 areas of fibrin were identified by the automated analysis, but these areas cover only 9.2% of the tissue area. The manual analysis of this image also labeled all background tissue as a combination of fluid and fibrin. However, simplifying this to 100% of tissue area, the sensitivity score based on area is 9.2%. For each of

these three cases, there is a large range in the sensitivity values measured using the two techniques, and neither is accurate.

There were an additional four image regions for which some specific areas of fibrin were identified in the manual analysis, along with all background tissue marked as fibrin. For regions [8001, 22001] and [14001, 20001] of slide D03-45740, all of the specifically marked areas of fibrin were found by the automated analysis, as well as numerous other areas of fibrin. Using the countable areas technique, sensitivity for both of these images is 100%. Using the comparison of areas technique, sensitivity scores for these two regions are only 11.2% and 22.8% respectively. For image region [32001, 20001] of slide D03-45740, five specific areas of fibrin were identified in the manual analysis. The automated analysis found only three of these specific areas, as well as 14 additional areas of fibrin, resulting in a sensitivity value of 85% based on countable areas. The 17 areas of fibrin marked by the automated analysis cover only 12.8% of the tissue area, however, resulting in a sensitivity based on area of 12.8%. Results for image region [8001, 42001] of slide D03-37303 were the least accurate. One specific area of fibrin was identified in the manual analysis, along with all unlabeled background tissue being identified as a combination of fluid and fibrin. This specific area of fibrin was not found, but one other area of fibrin was identified by the automated analysis, resulting in a sensitivity score of 50%. However, the one region of fibrin marked by the automated analysis comprises only 1.2% of the tissue area, resulting in a sensitivity score based on area of only 1.2%.

In general, sensitivity scores for fibrin detection were low, with values of less than 50% for 20 of the 36 images which contained fibrin. Especially noteworthy is the fact that seven image regions had 0% sensitivity. For two of these images, region [2001, 16001] of slide D96-4925 and region [22001, 26001] of slide D03-37303, the automated analysis did identify some fibrin within the image, but the identified regions did not correspond to the fibrin areas marked by the manual analysis. In the remaining five cases, no fibrin was detected by the automated analysis at all. This failure to find fibrin as a disease indicator in these images would lead to incorrect conclusions being drawn based on the automated analysis findings.

Low sensitivity values for fibrin detection are not entirely surprising, given the results of classifying regions in the training set of images. Once regions of interest are determined in the automated analysis, classification of areas as either fibrin or necrosis is based on the correlation value. Although use of this value provided extremely good results for necrosis, classification of

fibrin in the training set resulted in a sensitivity score of only 64%. One would expect that sensitivity scores for the slides in the analysis would not be significantly better than this value. Three slides, however, did achieve higher fibrin sensitivity scores: slide D03-00716 had an average sensitivity of 72.2% for its three image regions which contained fibrin; slide D03-45740 had an average sensitivity of 96.3% for its four image regions which contained fibrin; and slide D03-51213 had a sensitivity of 100% for its one region which contained fibrin. Only a single region of fibrin was present in the one region of D03-51213 which contained any fibrin, however, so this 100% result is not significant. It is also interesting to note that average sensitivity for fibrin for slide D03-10538 was only 34.7%, based on its five image regions which contained fibrin. Since all test regions in the training set were selected from this one slide, a higher sensitivity result would be expected.

As for specificity of fibrin detection, most images had few false positives, with an overall average of 3.1 FP per image. In cases where several areas of fibrin were incorrectly identified, many of these false positive areas tended to be located in regions of the image which had been marked “not to be analyzed”. This is the case for region [18001, 42001] of slide D03-45740, for example. In this case, 10 areas of fibrin were incorrectly identified. Eight of these areas were in a “not to be analyzed” area however, resulting in an Adjusted False Positive score for this image of two. Low false positive numbers led to good Positive Predictive Value values for many of the images, with fourteen images scoring more than 70%. There are also problems with the specificity of fibrin detection, however. Fourteen of the 50 images contained no fibrin, according to the manual analysis. The automated analysis correctly identified only one of these 14 as not containing fibrin. In another five of these images, the amount of fibrin marked by the automated analysis occupied less than 2% of the overall tissue area, and could be considered negligible. For the remaining eight cases, the total area of detected fibrin ranged from a low of 2.8% to a high of 28.1% of the total tissue area. For four of these eight images, including region [18001, 42001] of slide D03-45740 which had 28.1% of its tissue area identified as fibrin, all or almost all areas of fibrin are located within “not to be analyzed” regions of the tissue. Another significant contributing factor to false positives for fibrin is the misclassification of regions of necrosis as fibrin. For example, region [8001, 22001] of slide D03-00716 has four false positives. In this image, three areas identified as fibrin actually correspond to regions identified as necrosis in the manual analysis. The fourth area incorrectly identified as fibrin lies within a

region of tissue marked “not to be analyzed”. In total, 31 areas of necrosis in 13 images are incorrectly identified as fibrin, accounting for a third of all the false positives for fibrin. Mislabeling of fibrin as necrosis also occurs numerous times. An examination of results where detection of fibrin and necrosis are evaluated together is included in section 5.7.

5.6 Results of Necrosis Detection

5.6.1 Measurement of Results

There were no ambiguities when counting necrosis results. Unlike inflammatory cells and fibrin, there were no cases where the entire background tissue area was identified as containing necrosis. Instead, for necrosis all identified instances were marked as countable areas. The column “Percentage of Tissue Area” is also included in this Results chart (Table 5.18), however, since this data is informative when looking at some cases, such as those where necrosis was identified where none existed.

Table 5.18: Summary of necrosis detection results for all images

		Necrosis										
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*	Percentage of Tissue Area
D03-10538	12001, 20001	0	0	0	undef.		2	2	0.0%	0.0%		0.4
	12001, 36001	2	1	1	50.0%		4	4	20.0%	20.0%		14.6
	18001, 10001	4	1	3	25.0%		0	0	100.0%	100.0%		12.8
	18001, 40001	1	1	0	100.0%		1	1	50.0%	50.0%		7.5
	30001, 40001	1	1	0	100.0%		0	0	100.0%	100.0%		0.0
D03-00716	18001, 26001	0	0	0	undef.		0	0	undef.	undef.		0.0
	22001, 52001	2	0	2	0.0%		1	1	0.0%	0.0%		0.4
	14001, 24001	2	1	1	50.0%		0	0	100.0%	100.0%		3.3
	8001, 22001	11	0	11	0.0%		0	0	undef.	undef.		0.0
	34001, 28001	0	0	0	undef.		0	0	undef.	undef.		0.0
D03-45740	8001, 22001	5	2	3	40.0%		3	3	40.0%	40.0%		2.1
	18001, 42001	2	0	2	0.0%		7	1	0.0%	0.0%		7.5
	32001, 20001	3	0	3	0.0%		2	2	0.0%	0.0%		0.6
	14001, 20001	10	6	4	60.0%		4	4	60.0%	60.0%		10.0
	6001, 20001	7	2	5	28.6%		5	3	28.6%	40.0%		7.8
D03-11282	4001, 24001	1	0	1	0.0%		1	0	0.0%	undef.		1.0
	2001, 30001	7	3	4	42.9%		2	2	60.0%	60.0%		8.8
	24001, 46001	0	0	0	undef.		1	0	0.0%	undef.		0.6
	14001, 22001	1	0	1	0.0%		0	0	undef.	undef.		0.0
	26001, 34001	0	0	0	undef.		0	0	undef.	undef.		0.0
D96-4925	18001, 2001	0	0	0	undef.		0	0	undef.	undef.		0.0
	24001, 26001	0	0	0	undef.		0	0	undef.	undef.		0.1
	12001, 32001	0	0	0	undef.		0	0	undef.	undef.		0.0
	2001, 16001	0	0	0	undef.		0	0	undef.	undef.		0.0
	16001, 44001	0	0	0	undef.		0	0	undef.	undef.		0.0
D03-51216	6001, 20001	11	0	11	0.0%	**	0	0	undef.	undef.		0.0
	10001, 30001	0	0	0	undef.		2	2	0.0%	0.0%		0.9
	8001, 4001	6	0	6	0.0%		0	0	undef.	undef.		0.0
	8001, 28001	4	0	4	0.0%		5	4	0.0%	0.0%		2.6
	4001, 12001	10	0	10	0.0%		0	0	undef.	undef.		0.3
D03-37303	12001, 28001	0	0	0	undef.		1	1	0.0%	0.0%		0.2
	24001, 16001	4	0	4	0.0%		2	2	0.0%	0.0%		7.7
	22001, 26001	2	0	2	0.0%		0	0	undef.	undef.		0.0
	8001, 42001	0	0	0	undef.		4	3	0.0%	0.0%		20.4
	2001, 42001	0	0	0	undef.		17	17	0.0%	0.0%	??	19.4
D03-36878	1, 22001	0	0	0	undef.		0	0	undef.	undef.		0.0
	18001, 14001	0	0	0	undef.		1	1	0.0%	0.0%		0.2
	18001, 8001	0	0	0	undef.		0	0	undef.	undef.		0.0
	22001, 12001	0	0	0	undef.		10	10	0.0%	0.0%	??	4.1
	20001, 6001	0	0	0	undef.		0	0	undef.	undef.		0.0
N95-4825	20001, 10001	0	0	0	undef.		0	0	undef.	undef.		0.0
	10001, 40001	0	0	0	undef.		0	0	undef.	undef.		0.0
	22001, 24001	1	0	1	0.0%		0	0	undef.	undef.		0.0
	8001, 12001	0	0	0	undef.		1	1	0.0%	0.0%		0.3
	14001, 40001	0	0	0	undef.		1	0	0.0%	undef.		0.6
D03-51213	12001, 18001	0	0	0	undef.		0	0	undef.	undef.		0.0
	6001, 12001	0	0	0	undef.		6	1	0.0%	0.0%		6.4
	4001, 6001	0	0	0	undef.		2	0	0.0%	undef.		0.4
	10001, 18001	0	0	0	undef.		0	0	undef.	undef.		0.0
	16001, 16001	0	0	0	undef.		1	1	0.0%	0.0%		0.3
	AVERAGE				**22.6%**		**1.7**	**1.3**	**19.9%**	**23.8%**		

(For an explanation of rows indicated by a marker in the "*" column, refer to section 5.6.2)

5.6.2 Discussion of Results

Sensitivity results for the detection of necrosis were the lowest of all the detected disease indicators. Average sensitivity for the 22 of the 50 images which contained necrosis was only 22.6%. Sensitivity for 13 of these 22 images was 0%, and an additional four of these images had sensitivity scores below 50%. Specificity of necrosis detection, on the other hand, was very good. Of the 28 images which did not contain any necrosis, no necrosis was indicated in the automated analysis either for 15 of these images, and only a negligible amount of necrosis was indicated in an additional nine of them.

Since the training images for fibrin and necrosis detection were taken from slide D03-10538, one would expect that better results might be obtained for image regions of this slide than for images of other slides. Although this was not found to be the case for fibrin, slide D03-10538 did have significantly better sensitivity scores for necrosis than all other slides. Four of the five image regions of slide D03-10538 contained necrosis, and sensitivity scores for these four regions were 50%, 25%, 100% and 100%, for an average sensitivity for this slide of 68.75%. Average sensitivities for the image regions of each of the other slides produced a maximum value of only 25.7% for any one slide, and a minimum value of 0%. Furthermore, slide D03-10538 is the only one of the seven slides which contained necrosis to record no 0% sensitivity scores for any of its image regions.

The failure of the algorithms to correctly identify necrosis is most clearly evident on slide D03-51216, where no necrosis was found by the automated analysis in any of the four of its five image regions which contained necrosis. A total of 31 areas of necrosis across these four image regions were missed.

One of the main reasons for the poor sensitivity scores for detection of necrosis was misclassification of necrotic areas as fibrin. For example, for image region with upper left co-ordinates [8001, 28001] of slide D03-51216, four regions of necrosis were identified by the manual analysis. None of these four regions was marked as necrosis by the automated analysis, resulting in a 0% sensitivity score, but all four of them were detected and labeled as fibrin by this analysis. Mislabeling of areas of necrosis as fibrin occurred in all three other image regions of slide D03-51216 which also had 0% sensitivity for necrosis. For region [6001, 20001], six of the 11 missed areas of necrosis were identified as fibrin; for region [8001, 4001], three of the six missed areas were identified as fibrin; and, for region [4001, 12001], four of the 10 missed areas

were identified as fibrin. If these mislabeled necrotic regions were considered detected, average sensitivity for the four regions of this slide would rise to 61% from its current average of 0%.

This same misclassification problem occurs with the other images for which sensitivity of necrosis detection was 0%. Of the 13 image regions with 0% sensitivity, mislabeling of areas of necrosis as fibrin occurs in eight of them. Across all images, a total of 32 regions of necrosis which were counted as false negatives were actually correctly detected but mislabeled as fibrin. This seems to indicate that the problem with fibrin and necrosis detection in the automated analysis may be primarily in the final step of the algorithm, where detected regions of interest are classified as fibrin, necrosis or "other". For this reason, an examination of the combined sensitivity of this algorithm in detecting regions which are either fibrin or necrosis is included in section 5.7.

Although sensitivity results for necrosis detection are disappointing, specificity values for detection of this feature are promising. There are 28 image regions which, according to the manual analysis, contain no necrosis. Of these, no necrosis is found by the automated analysis in 15 of them. Some amount of necrosis is indicated in the other 13 image regions where none should be found, but for nine of these the number of areas of necrosis marked by the automated analysis is two or fewer, and the total amount of tissue area labeled as necrosis is less than 1% of the total tissue area. Therefore, the specificity of necrosis detection is satisfactory for 24 of these 28 image regions.

Specificity of necrosis detection is also satisfactory when looking at all image regions, whether or not they contained necrosis according to the manual analysis. Across all 50 image regions, the average number of false positives was only 1.7 per image. The Positive Predictive Value (PPV) score at 19.9%, and Adjusted Positive Predictive Value at 23.8%, are deceivingly low, however. There were 28 image regions in which any necrosis was found by the automated analysis, and the average PPV is the obtained from these 28 values. Nineteen of these 28 PPV values were 0%, indicating that all regions of necrosis found by the automated analysis were incorrect. Of these 19 image regions, 12 had a total of 1% or less of the total tissue area marked as necrosis, which is a negligible amount.

The four image regions with poor specificity were regions [8001, 42001] and [2001, 42001] of slide D03-37303, region [22001, 12001] of slide D03-36878, and region [6001, 12001] of slide D03-51213. The two regions of slide D03-37303 were the worst cases. The

automated analysis of region [8001, 42001] of slide D03-37303 identified a total of four areas of necrosis, covering a total of 20.4% of the tissue area. For region [2001, 42001] of the same slide, 17 areas of necrosis were marked, covering a total of 19.4% of the tissue area. In both cases, the problem again appears to stem from misclassification of fibrin as necrosis. For both of these regions of slide D03-37303, the entire area of the tissue was marked as fluid and fibrin in the manual analysis. This means that all identified areas of necrosis in both images would have been correct if they had been classified as fibrin instead. This same problem is even more evident in image region [22001, 12001] of slide D03-36878. In this case, 10 areas of necrosis covering a total of 4.1% of the tissue area were incorrectly identified by the automated analysis. Eight of these 10 areas correspond to regions which were specifically marked as fibrin in the manual analysis. Region [6001, 12001] of slide D03-51213 is the exception. Here, six areas of necrosis covering a total of 6.4% of the tissue area were identified. Five of the six are located in an area of the slide which was marked "not to be analyzed" in the manual analysis.

5.7 Examination of Combined Fibrin and Necrosis Results

5.7.1 Measurement of Results

Misclassification of fibrin as necrosis, and vice versa, occurred numerous times in the automated analysis. The most common misclassification was labeling areas of necrosis as fibrin, contributing both to high false positive counts for fibrin and low true positive (low sensitivity) counts for necrosis. These mislabeling errors indicate that the most significant problem in fibrin and necrosis detection algorithms lies in the final step, where detected regions of interest are classified as fibrin, necrosis, or "other". The following results chart is included in order to examine the effect of these misclassification errors. In this chart, results for fibrin and necrosis detection are combined, such that for any area identified as fibrin or necrosis in the manual analysis, if the area is found and labeled as either fibrin or necrosis by the automated analysis it is considered a true positive. The figures shown in the "Number of Areas" column of this chart are the sums of the areas of fibrin and necrosis identified by the manual analysis. For the seven cases in which all unlabelled background tissue was identified as fibrin in the manual analysis, these figures list the number of specific areas of fibrin and necrosis identified in these images, and include a "+++" symbol to indicate that more area of tissue is also considered fibrin or

necrosis. For each of these cases, any areas of fibrin or necrosis identified by the automated analysis which correspond to the nonspecifically labeled background fibrin area in the manual analysis are also counted as true positives, consistent with the handling of these areas in the assessment of fibrin results. Consequently, the total of true positives and false negatives may be larger than the figure listed in the “Number of Areas” column for each of these images.

Table 5.19: Summary of detection results for fibrin and necrosis together

		Fibrin and Necrosis										
Slide	Region	Number of Areas	TP	FN	Sensitivity	*	FP	Adjusted FP	Positive Predictive Value	Adjusted Positive Predictive Value	*	Percentage of Tissue Area
	12001, 20001	25	3	22	12.0%		2	2	60.0%	60.0%		8.7
	12001, 36001	25	13	12	52.0%		1	1	92.9%	92.9%		41.5
D03-10538	18001, 10001	7	2	5	28.6%		4	0	33.3%	100.0%		30.3
	18001, 40001	23	14	9	60.9%		0	0	100.0%	100.0%		19.0
	30001, 40001	24	6	18	25.0%		1	0	85.7%	100.0%		7.1
	18001, 26001	3	2	1	66.7%		2	2	50.0%	50.0%		2.3
	22001, 52001	2	0	2	0.0%		7	7	0.0%	0.0%		3.5
D03-00716	14001, 24001	6	4	2	66.7%		0	0	100.0%	100.0%		8.5
	8001, 22001	11+++	5	8	38.5%	*	1	0	83.3%	100.0%		6.4
	34001, 28001	0	0	0	undef.		1	1	0.0%	0.0%		0.5
	8001, 22001	7+++	14	2	87.5%	*	3	0	82.4%	100.0%		13.3
	18001, 42001	2	1	1	50.0%		17	5	5.6%	16.7%		35.6
D03-45740	32001, 20001	8+++	22	3	88.0%	*	0	0	100.0%	100.0%		13.4
	14001, 20001	12+++	24	1	96.0%	*	0	0	100.0%	100.0%		32.8
	6001, 20001	7+++	6	5	54.5%	*	14	0	30.0%	100.0%		29.3
	4001, 24001	16	9	7	56.3%		8	1	52.9%	90.0%		14.6
	2001, 30001	24	11	13	45.8%		3	1	78.6%	91.7%		38.8
D03-11282	24001, 46001	0	0	0	undef.		5	3	0.0%	0.0%		13.7
	14001, 22001	3	0	3	0.0%		0	0	undef.	undef.		0.6
	26001, 34001	0	0	0	undef.		4	4	0.0%	0.0%		2.8
	18001, 2001	3	0	3	0.0%		0	0	undef.	undef.		0.0
	24001, 26001	1	1	0	100.0%		5	2	16.7%	33.3%		8.3
D96-4925	12001, 32001	16	0	16	0.0%		0	0	undef.	undef.		0.0
	2001, 16001	2	0	2	0.0%		2	2	0.0%	0.0%		0.4
	16001, 44001	0	0	0	undef.		1	1	0.0%	0.0%		0.3
	6001, 20001	25	17	8	68.0%		3	1	85.0%	94.4%		11.5
	10001, 30001	23	8	15	34.8%		1	0	88.9%	100.0%		15.7
D03-51216	8001, 4001	21	7	14	33.3%		2	2	77.8%	77.8%		17.3
	8001, 28001	27	21	6	77.8%		2	1	91.3%	95.5%		25.3
	4001, 12001	33	14	19	42.4%		2	2	87.5%	87.5%		8.2
	12001, 28001	15	1	14	6.7%		8	6	11.1%	14.3%		3.7
	24001, 16001	9	1	8	11.1%		0	0	100.0%	100.0%		15.5
D03-37303	22001, 26001	10	1	9	10.0%		0	0	100.0%	100.0%		0.4
	8001, 42001	1+++	4	1	80.0%	*	1	0	80.0%	100.0%		21.6
	2001, 42001	many	27	0	100.0%	*	0	0	100.0%	100.0%		28.6
	1, 22001	11	1	10	9.1%		1	1	50.0%	50.0%		0.9
	18001, 14001	3	2	1	66.7%		3	3	40.0%	40.0%		2.5
D03-36878	18001, 8001	0	0	0	undef.		0	0	undef.	undef.		0.2
	22001, 12001	7	7	0	100.0%		9	9	43.8%	43.8%		16.8
	20001, 6001	0	0	0	undef.		4	4	0.0%	0.0%		1.5
	20001, 10001	2	0	2	0.0%		0	0	undef.	undef.		0.0
	10001, 40001	0	0	0	undef.		2	1	0.0%	0.0%		1.6
N95-4825	22001, 24001	16	0	16	0.0%		0	0	undef.	undef.		0.0
	8001, 12001	12	4	8	33.3%		3	3	57.1%	57.1%		2.7
	14001, 40001	0	0	0	undef.		8	1	0.0%	0.0%		4.1
	12001, 18001	0	0	0	undef.		3	1	0.0%	0.0%		4.2
	6001, 12001	0	0	0	undef.		12	2	0.0%	0.0%		14.3
D03-51213	4001, 6001	0	0	0	undef.		5	2	0.0%	0.0%		1.3
	10001, 18001	1	1	0	100.0%		3	2	25.0%	33.3%		2.6
	16001, 16001	0	0	0	undef.		7	7	0.0%	0.0%		3.1
	AVERAGE				**44.8%**		**3.2**	**1.6**	**47.9%**	**55.2%**		

5.7.2 Discussion of Results

The 44.8% sensitivity score for detection of combined fibrin and necrosis areas is a large improvement over the low 22.6% sensitivity for necrosis alone, but is slightly worse than the 45.3% sensitivity for fibrin alone. The most significant contributors to this drop in sensitivity score over fibrin alone are the cases where all background tissue was nonspecifically labeled as fibrin in the manual analysis. For region [8001, 22001] of slide D03-00716, for example, although only two areas of fibrin were identified by the automated analysis, this was considered 100% sensitivity for fibrin detection based on the countable regions technique. When looking at the combined results for this image, three of the 11 areas of necrosis marked in the manual analysis were detected and identified as fibrin, in addition to the two correctly identified areas of fibrin, resulting in a true positive score of five. Eight of the 11 areas of necrosis were not detected, however, resulting in a sensitivity score for this image of only 38.5%. This is an improvement over the 0% sensitivity scored for this image for necrosis detection alone, but is much lower than the 100% sensitivity recorded for fibrin detection alone. Similarly, region [6001, 20001] of slide D03-45740 had a much lower sensitivity score in the combined results than for fibrin alone. In this case, only three areas of fibrin were identified by the automated analysis, but based on the countable regions technique this resulted in a 100% sensitivity score. For this image, only two of the seven specifically marked areas of necrosis were found. In addition, one area of necrosis marked by the automated analysis corresponded to the nonspecifically marked background fibrin area, and was counted as a true positive in the combined results, summing to a total of six true positives. With five areas of necrosis not detected, the resulting sensitivity in the combined results is 54.5%, an improvement over the 28.6% score for necrosis alone, but lower than the 100% score for fibrin alone. The total percentage of tissue area identified as either fibrin or necrosis in this image by the automated analysis is only 29.3%, but this would represent an improvement in sensitivity score based on area for fibrin detection.

On the specificity side, the 47.9% Positive Predictive Value and 55.2% Adjusted Positive Predictive Value represent a significant improvement over the values for both fibrin and necrosis detection alone, reflecting many cases where regions of necrosis which were identified as fibrin, or vice versa, are no longer considered false positives.

Some of the most significant score improvements when fibrin and necrosis results are combined include region [8001, 28001] of slide D03-51216, region [2001, 30001] of slide D03-11282, region [14001, 20001] of slide D03-45740, and region [22001, 12001] of slide D03-36878, all of which had reductions in false positive and false negative scores of nine or more. For image with upper left co-ordinates [8001, 28001] of slide D03-51216, for example, the true positive count increased by eight, the false negative count dropped by eight, and the false positive count dropped by six in the combined results compared to the individual results. In this image, four areas of fibrin marked in the manual analysis were found by the automated analysis but identified as necrosis (Figure 5.38). In the individual results, these areas were counted as four false negatives for fibrin and four false positives for necrosis. In the combined results these areas are scored instead as four true positives. In addition, four areas of necrosis marked by the manual analysis were identified as fibrin in the automated analysis. These four false negatives for necrosis and two false positives for fibrin end up as four true positives in the combined results. (There is only a reduction of two in the false positive score for these particular areas since in two cases the areas of necrosis fell within larger areas identified as fibrin by the automated analysis. These large areas of fibrin also happened to overlap with valid areas of fibrin, and so were counted as true positives for fibrin in the individual results).

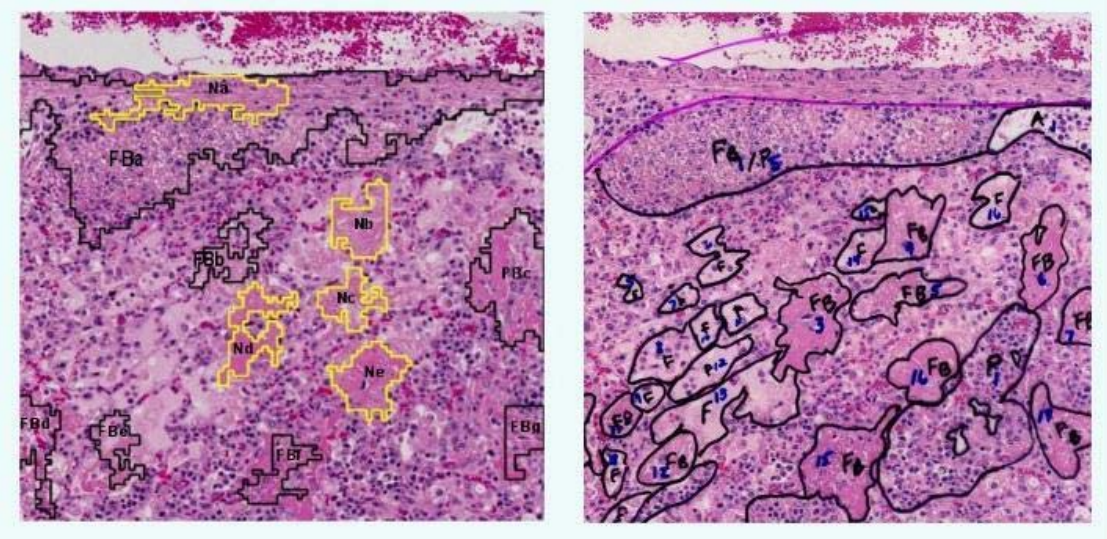


Figure 5.38: Part of image region [8001, 28001] of slide D03-51216. The automated assessment result for fibrin and necrosis only is shown on the left. The manual assessment result for all disease indicators is shown on the right. Fibrin regions marked as FB3, FB4, FB5, and FB16 in the manual assessment were detected in the automated analysis, but identified as regions of necrosis labeled Nd, Nb, Nc, and Ne respectively.

Not detecting necrosis in images where it exists was a large problem in the necrosis detection results. Mislabeling of necrosis as fibrin was one of the main reasons for this failure. By combining the necrosis and fibrin results, there is a significant improvement in the number of cases in which the feature is marked as absent when it actually does appear, but the problem is not eliminated entirely. In the individual results, necrosis detection failed to find any necrosis in 13 images where it should have been found. In the combined results, some fibrin or necrosis is found in 10 of these 13 images. There are still seven images which contain either fibrin or necrosis, however, in which none is indicated when the fibrin and necrosis results are combined.

5.8 Problems with the Method of Scoring Results

As has been discussed throughout this chapter, scoring results of the automated analysis using sensitivity and positive predictive value numbers is problematic. These values are greatly affected by the way in which regions of disease indicators are marked, both in the automated analysis and in the manual assessment. For example, suppose an image contains a single area of necrosis. One automated analysis result locates this area, and also incorrectly identifies a second region of the image as necrosis, resulting in a positive predictive value of 50%. A second automated analysis identifies the exact same two areas: the correctly located region of necrosis, and the false positive for necrosis. In the second analysis, however, the false positive area is marked as two separate regions which together occupy the same area as the one false positive in the first analysis. Although the results of both automated analyses identify exactly the same areas, the false positive count for the second analysis is higher, and positive predictive value for this analysis is only 33%. Although both analyses should be judged as being of equal accuracy, PPV numbers indicate that the first analysis is more accurate than the second.

The huge impact that a subtle change in identifying areas within an image can have on the result scores is clearly shown in the example below. In this case, one image was assessed by the same pathologist on two separate occasions. Without intentionally changing the analysis, the pathologist used two different approaches when analyzing this image. In the first analysis, very specific areas were marked, resulting in a total of 15 areas of fibrin identified in the image (Figure 5.39). In the second analysis of the same image, the pathologist used a less specific approach, identifying fewer but larger areas of fibrin. In this case, only three areas of fibrin were

identified in the image (Figure 5.40). Both approaches are acceptable and correct from the pathology perspective. However, the two different manual analyses result in vastly different scores for the automated assessment of this image. The thesis software always attempts to split out the most precise regions of disease indicators possible. In this case, the automated analysis identifies 15 areas of fibrin within this image (Figure 5.41).

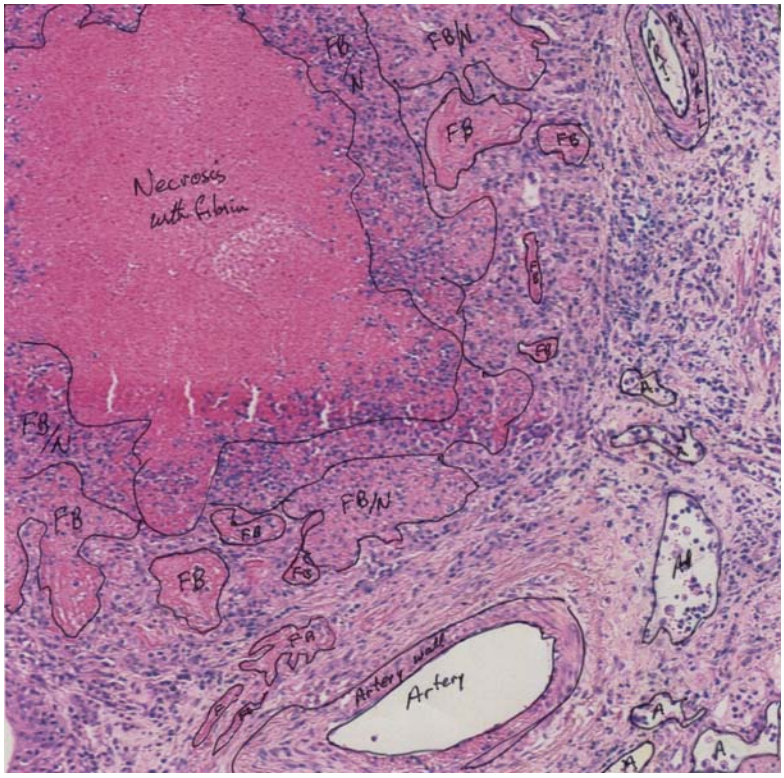


Figure 5.39: First expert manual analysis of an image

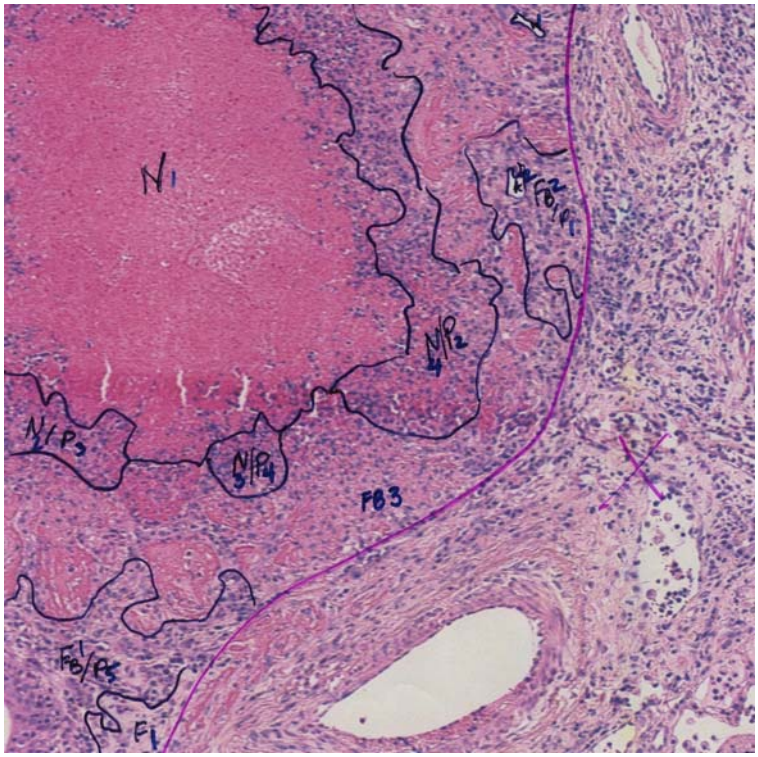


Figure 5.40: Second manual analysis of the same image by the same expert

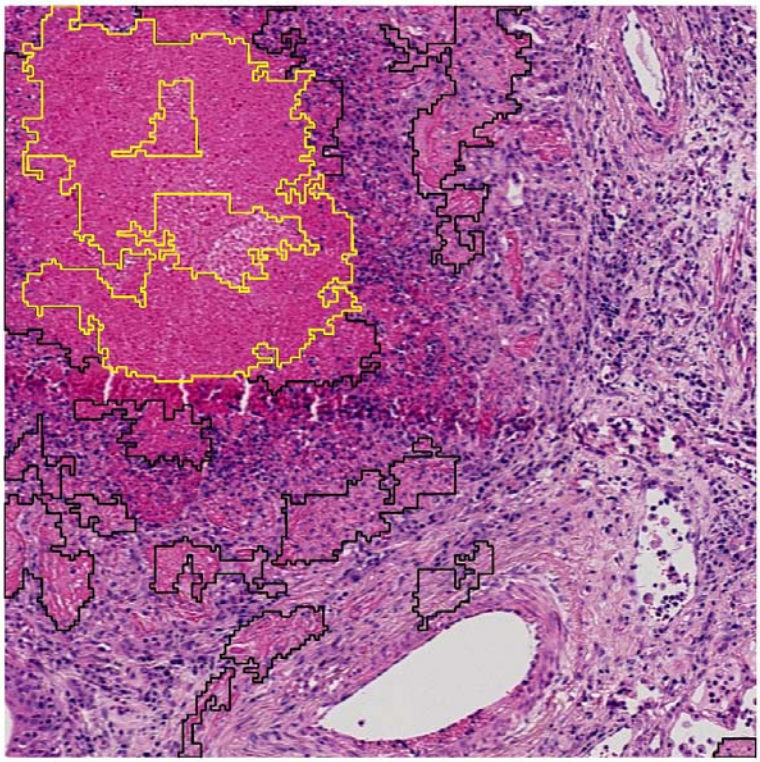


Figure 5.41: Automated assessment of areas of fibrin (marked in black) and necrosis (marked in yellow)

When results for fibrin are tallied by comparing to the first manual analysis, 11 of the 15 areas of fibrin are found, resulting in TP score of 11 and FN score of 4, for a sensitivity of 73%. (It should also be noted that there is a very good match of region boundaries between the automated and manual analysis for some of the fibrin areas). The false positive score for fibrin in this case is 2, resulting in a positive predictive value of 85%.

When the same automated assessment is evaluated by comparison to the second manual analysis, the results are much worse. In this case, one of the large areas of fibrin marked in the manual analysis is found; this one area overlaps with six areas of fibrin marked in the automated assessment. The remaining two areas of fibrin in the manual analysis are not found, resulting in a TP score of 1 and FN score of 2, for a sensitivity of only 33%. Nine areas of fibrin marked in the automated assessment have no corresponding areas in this manual analysis, resulting in a positive predictive value of only 10%.

This one example clearly illustrates how arbitrary the sensitivity and positive predictive value numbers can be. When compared to the first expert manual analysis, the software results are very good. When the same software output is compared to the second expert analysis, the results appear to be extremely poor.

CHAPTER 6
CONCLUSIONS, DISCUSSION AND FUTURE WORK

The main objective of this thesis project was to develop software which, by employing appropriate image processing techniques, could identify and measure fluid-filled alveoli, areas of inflammatory cells, fibrin, and necrosis within images of cattle lung tissue samples. The image analysis techniques implemented in the thesis software all show potential. Detection of each of the four disease indicators by the software was successful to some extent. Furthermore, for each disease indicator, results for some images within the test set were extremely good. Nevertheless, when results of the entire set of test cases are examined, the overall results of the software analysis are not accurate enough that a diagnosis of disease states could be made with any amount of certainty, if the diagnosis was based only on the information provided by the software.

In the case of detection of fluid-filled alveoli, the software demonstrates excellent sensitivity and accuracy at finding alveoli within the images, with an overall sensitivity score of 94.8% and average positive predictive value of 83.2%. Finding alveoli is only the first step, not the goal however, in this feature detection routine. When assessing alveolar contents to determine the prevalence of fluid, the accuracy drops. Sensitivity of detection of fluid-filled alveoli is still very good, with a score of 89.8%. The issue is that many more alveoli are judged to contain fluid by the automated analysis than are categorized this way in the expert manual analysis. Within the set of alveoli, a positive predictive value of only 41.9% indicates that more of the alveoli which are labeled as fluid-filled are incorrect than correct. Specificity of identifying alveolar contents shows that only 12% of the air-filled alveoli are correctly identified by the automated analysis as containing air. Use of the software to determine whether the presence of fluid is a factor in the condition of the lung sample would lead to over-diagnosis of fluid-related diseases.

Presence of inflammatory cells also tends to be over-predicted by the software. Measuring the accuracy of detection of inflammatory cells using sensitivity and positive predictive values is difficult due to the fact that these areas do not have clearly defined boundaries. As a result, counting specific areas of inflammatory cells is challenging, and

arguably of little value. However, even if the specific assessment numbers are not considered, the detection results do show that areas of inflammatory cells are found by the automated analysis in 13 cases where none should be, including in all five image regions of one slide. Although the percentage of tissue area identified as containing these cells is relatively low in some of the cases, this false identification could lead to incorrect diagnoses. Also, for many of the images in which areas of inflammatory cells did exist, there were significant numbers of additional areas which were incorrectly labeled as containing these cells by the automated analysis, potentially over-emphasizing the importance of this indicator in the disease state of those samples. It is difficult to determine whether these incorrectly identified areas would result in incorrect conclusions being drawn, however. In the case of inflammatory cells, the most useful information produced by the automated analysis is likely to be the percentage of tissue area which contains these cells. The number of specific areas of inflammatory cells is likely to be of little use, and would probably not be utilized in diagnosing the disease present, so the fact that there are false positives for most of the images may not be relevant. The accuracy of inflammatory cell detection should have been evaluated by comparing the amount of tissue area identified as infiltrated by neutrophils and/or macrophages by the manual and the automated analyses. Without this comparison, it is difficult to conclude whether the automated analysis results are helpful or misleading in arriving at a diagnosis. The positive aspect to detection of inflammatory cells is the fact that the automated analysis does find some amount of these cells in all but one of the images in which they are present.

With a sensitivity score of 45.3% and a positive predictive value of 41.1%, less than half the existing areas of fibrin are found by the thesis software and less than half of the areas identified by this software as fibrin actually are. In addition, no fibrin was detected in five images which contained fibrin, and more than a negligible amount of fibrin was falsely identified in eight images which contained none. At this stage, using the software to determine whether fibrin is a factor in the disease states of the samples would be unreliable.

Necrosis results are worse still, with a sensitivity score of only 22.6% and a positive predictive value of 19.9%. However, the automated analysis is reasonably accurate at identifying samples which do not contain necrosis. Of the 28 images where necrosis was not present, the software found either no necrosis or only a negligible amount of necrosis in 24 of them. The low positive predictive value is deceiving, since in almost all of the cases where PPV

was 0% for an image, only a negligible amount of necrosis was identified by the automated analysis. If these cases are not included in the average, PPV would be 34.9%. It is worth noting that a major cause of the low sensitivity score for necrosis detection was mislabeling of necrotic areas as fibrin. When detected fibrin and necrosis results are counted together, the improvement in the sensitivity score to 44.8% and in positive predictive value score to 47.9% illustrates the magnitude of this misidentification.

Without making drastic changes to the image processing techniques used in the thesis software, there are a number of areas in which some improvements to the accuracy of the detection algorithms could be made. There are also numerous alternate image processing techniques which may be useful in detecting these disease-indicating features, and which are worth further investigation. In addition, if detection algorithms were improved to obtain a satisfactory level of accuracy, there is additional work which would then need to be done in order to transform the programs into a useful tool for pathologists. Each of these areas of future work is described below.

6.1 Potential Improvements to Existing Algorithms

Although there are a number of causes of the over-prediction of fluid-filled alveoli by the software, the main contributor to this problem may in fact be an incorrect assumption about how alveolar contents should be classified. Any alveolus which the programs found to contain more than 33% pink pixels was identified as fluid-filled in the automated analysis. In actual pathological assessments, there is no defined percentage of fluid which all pathologists use to distinguish a fluid-filled alveolus from one which contains air. In fact, an assessment of the significance of the presence of fluid is more likely to be made from the pathologist's overall impression of the tissue rather than counting the number of individual alveoli which contain fluid or do not. The 33% cutoff used in the software was chosen somewhat arbitrarily, and was then applied to both the automated and manual analyses of this project. Since the programs can only assess fluid content numerically, such a threshold is required. However, the validity of this particular threshold value should be evaluated by having a pathologist label each alveolus in a set of test images as fluid-filled or not based on whatever criteria would normally be used to make this distinction. The actual percentages of pink pixels in each of the alveoli which were considered fluid-filled should then be measured to determine an appropriate threshold value.

Although the specificity of labeling alveoli as fluid-filled is the largest area for improvement in the alveoli detection algorithms, identification of alveolar boundaries could also be improved. In some cases, areas of fluid which were stained a very dark pink colour were not included in alveolar regions, but instead considered part of the alveolar walls. Improvements to the accuracy of alveolar boundaries could potentially be achieved by adding a region-growing step to this algorithm. In this additional step, once alveolar boundaries are determined using the current RGB classification, pixels surrounding each alveolus would be examined and merged into the alveolar region if a homogeneity criterion is satisfied. The homogeneity measure could be based on colour, texture, or both. Lezoray and Cardot [21], for example, describe the use of colour watershed as such a region-growing technique to improve the initial colour pixel classification results in their assessment of cytology images.

The current simplistic method in which areas along the edges of the images are determined to be either alveoli or part of the slide background should also be improved. As was discussed in section 4.1, the current technique will incorrectly assume that any alveolus which happens to lie at the corner of an image is actually part of the background, and this alveolus will not be evaluated by the programs. Although this does not contribute in any significant way to the alveoli assessment results, it is a known problem which should be corrected.

As for detection of inflammatory cells, before any modifications to the programs are undertaken, the first step needs to be an improved evaluation of the current detection results. Assessing the accuracy of inflammatory cells detection by counting areas marked by the automated and manual analyses is not adequate. Unlike alveoli, regions of inflammatory cells do not necessarily have distinguishable boundaries, so identification of distinct areas is difficult. Detection results should be evaluated based on the amount of area in each image which is identified as containing inflammatory cells. This area information is currently available for the automated assessment, but needs to be measured for the manual analysis also.

A more rigorous selection of all internal parameters used in the inflammation detection algorithms may improve overall results of these programs. For inflammation detection, four parameters are used. The choice of six as the distance to be used with the gradient function was based on the size of leukocyte nuclei within the images. The three remaining parameters (threshold of gradient function considered a significant edge magnitude, minimum edge density

based on number of edge pixels within a neighbourhood, minimum size of areas in edge density image) were all determined experimentally, based on a very limited set of test images. Re-evaluation of appropriate values for these parameters based on a more complete set of test cases would be beneficial.

Selection of internal parameters could also be improved in the detection of fibrin and necrosis, and inappropriate values for some of these parameters are likely to be the cause of some inaccuracies with these algorithms. Internal parameters used in the image segmentation part of these programs, such as weightings used for colour distribution and edge patterns in the homogeneity measure, for example, could potentially be improved if thorough experimentation using a larger set of test cases was conducted. However, the segmentation results are generally good, and improvements in this area would not have the most significant impact on analysis results. Focus on improving detection of fibrin and necrosis should instead be placed on classification of regions after the initial segmentation of image regions is complete. The set of image regions from the output of the image segmentation step which are then evaluated to see if they might be fibrin or necrosis are initially selected by the percentage of pixels which were classified as "pink" in the RGB clustering step. The threshold of 57% pink pixels was determined experimentally using a test set of 17 regions from two different images. The purpose of this threshold is merely to reduce the number of regions which then have to be analyzed further, so inclusion of additional areas is acceptable provided that potential areas of fibrin or necrosis are not eliminated. Evaluation of the 50 images used to assess the thesis program results showed that there was more variance in the appearance of regions of fibrin, and also in regions of necrosis, than was originally thought. A larger set of test cases, including more regions of fibrin and necrosis from various images, should be used to determine whether this 57% threshold is too restrictive.

Evaluation of the manual analysis results also revealed a misunderstanding in the creation of the fibrin and necrosis detection algorithms. The thesis programs classify regions of interest into exclusive categories fibrin, necrosis, or "other". In the manual assessment, however, 16 regions in 11 images were marked as containing both fibrin and necrosis. This combination of contents was not considered in the programs, and more analysis is required to determine how the classification step needs to be altered to accommodate this combination.

As was previously discussed, a number of areas of necrosis were mislabeled as fibrin, and vice versa. In these cases, the initial image segmentation and selection of regions of interest were successful, but the final classification step failed. It is also likely that there are a number of areas of fibrin or necrosis which were segmented correctly but misclassified as "other", contributing to the low sensitivity scores for these features. Classification based on the four co-occurrence matrix textural features provided by Matlab (contrast, correlation, homogeneity, energy), and classification based on correlation alone, were both evaluated. Results of classification of the set of fibrin and necrosis test cases were not particularly good with either classification test. When using the set of four textural features, sensitivity of 83% and specificity of 96% was achieved for necrosis detection, but for fibrin these values were only 45% sensitivity and 67% specificity, and for "other" areas the values were 60% sensitivity and 71% specificity. When just correlation was used, necrosis results were 100% sensitivity and 100% specificity, fibrin results were 64% sensitivity and 57% specificity, and "other" results were 40% sensitivity and 76% specificity. Classification based on correlation value alone was used to produce the automated assessment results. Necrosis results are very good when either classification function was used on the test cases, which makes the fact that necrosis results for the 50 images of the full assessment are so low even more interesting. This shows that the set of necrosis test cases was not adequate. In fact, during the assessment of the 50 images, much more variation in the appearance of areas of both necrosis and fibrin was observed than was evident in the test cases. Given the very small size of the set of test cases (containing only 11 areas of fibrin, and six areas of necrosis), it is not surprising that the test set did not cover the full range of different areas which should be classified as fibrin or necrosis. Also, all test cases were selected from two different images areas of a single slide. This too is inadequate. A more comprehensive set of test cases should be created, using example regions from several different slides, and the classifier should be retrained based on these cases. Even before the set of test cases are expanded, it would be interesting to compare the results of re-classifying regions all 50 images using the classifier based on the four textural features. Although necrosis results were slightly worse using this classifier compared to use of correlation value alone, sensitivity of fibrin detection and specificity of "other" detection were better.

When the set of test cases is expanded to include a more representative set of areas, it may be the case that a single co-occurrence matrix textural feature can successfully separate

fibrin, necrosis, and "other" areas. Correlation was the single feature which had the best results when used with the existing set of test cases. For necrotic areas, correlation had a unique range of values when compared to fibrin and "other" areas. However, there was some amount of overlap of the ranges of correlation values for fibrin and "other". Given this outcome with the test cases, one would have predicted a lower number of false positives for necrosis than for fibrin. This did turn out to be the case with the full assessment, with the number of false positives per image averaging only 1.7 for necrosis, compared to 3.1 per image for fibrin. Positive predictive value for necrosis was much lower than that for fibrin, however. As was previously discussed, this low PPV value is due to the large number of cases in which true positives (and consequently also PPV) for necrosis were 0.

If classification is to be based on textural features computed from co-occurrence matrices, it may also be worth examining some of the other features available. The four texture features (contrast, correlation, homogeneity, and energy) included in the classifier of this project are four of the five most commonly used features [13], but there are a total of fourteen texture features calculated from co-occurrence matrices which are described in the original paper by Haralick et al. [3]. It would be worth investigating whether a classifier which included additional textural features produced better results.

Distances selected for computing the co-occurrence matrices should also be re-examined. For the thesis software, co-occurrence matrices were constructed using both a distance of one and a distance of three. For the final classification of regions in this project, correlation values from co-occurrence matrices using distance one were used. Both selected distances were chosen arbitrarily. Although there is no established technique for selecting appropriate distances to use, it may be worth experimenting with different distance values to determine which provides the best results.

Also, although overall assessment numbers would not be improved through this effort, the final step of the Ojala-Chen method should be implemented. Pixelwise classification is used as the last step in this technique in order to improve region boundaries. This step was not included in the thesis software due primarily to time constraints. Adding this step would smooth otherwise jagged boundaries of detected areas of fibrin and necrosis.

As an additional note, in the thesis software the Ojala-Chen method was used only to segment image regions for the purposes of finding areas of fibrin and necrosis. The output of the

Ojala-Chen method is an image which has been segmented into regions which are judged to be homogenous in terms of colour and texture. There is nothing about this segmentation which is particular to fibrin and necrosis detection. This image segmentation could be used to detect any image features, including alveoli and inflammation. Highlighted images produced by the Ojala-Chen algorithm do show promising results in outlining alveolar regions, for example. More investigation would be required to find the appropriate additional processing to determine whether the detected regions are air-filled alveolar space, fluid, inflammation etc., but use of the Ojala-Chen method as a single means of isolating all regions of interest may be a possibility.

6.2 Alternate Processing Techniques

Any number of image processing techniques could have been used as part of this project. The set of techniques implemented all show potential, and with additional work as described in the previous section, results could possibly be improved. Other alternate image processing techniques could also be explored. A few algorithms which may be useful in this type of image processing problem are described below.

In the alveoli detection step, RGB classification was used to identify pixels within alveolar areas. A similar classification based on the HSV colourspace may be more useful, and more applicable to this particular problem. With histology slides, stains used to prepare samples are basically monochromatic, which would be reflected in little variance in the hue of the HSV system. The information provided by staining tissues is conveyed in the intensity of the staining, and would be evident in the HSV colourspace by variations in saturation and value numbers. HSV colourspace was employed by Tanaka et al. [18] in their classification of tumor cells in microscopic images, and also by Pavlova et al. [19] in their analysis of microscopic images of biological samples. This technique is worth exploring for the analysis of the lung tissue samples of this problem domain also.

Alternate methods of evaluating texture variations are also worth exploring. For the thesis software, texture features calculated from co-occurrence matrices were used, but this is just one of many possible techniques. Law's Texture Energy Measures would be one possible alternate technique. These measures provide a set of vector filters which respond to different texture features when applied to an image. A convolution mask can be created from the vectors corresponding to the texture features of interest (spots and edges, for example). Applying the

convolution mask to an image provides statistics which describe the texture. Use of Law's Texture Energy Measures as a means of classifying regions of interest detected by the Ojala-Chen algorithm already implemented in the thesis software is one possibility. Utilizing other texture measures provided by Matlab is another option, with the advantage that limited programming is required. Matlab's Image Processing Toolbox, version 5 and later, provides three texture filter functions: range, entropy, and standard deviation. Each function defines a neighbourhood around a pixel of interest and calculates the texture statistic for this neighbourhood. Exploration of these features may be worthwhile, particularly since minimal effort would be required to test the usefulness of these texture statistics.

6.3 Usability Improvements

Currently, the thesis software is executed by running three separate scripts. The first script (*analyzeTissue.m*) runs all programs involved in alveoli and inflammation detection. Highlighted images and an image analysis report are created and saved. The next script (*OjalaChen.m*) runs the hierarchical splitting and agglomerative merging steps of the fibrin and necrosis detection algorithm, and stores an image and corresponding data in which all homogenous areas within the image are identified. The final script (*findCooccurence.m*) calculates co-occurrence matrices and identifies areas which appear to be fibrin or necrosis. Highlighted images and an analysis report are also saved as output of this script. The current set of scripts requires that the image area of interest (a 2000 x 2000 pixel image region, for the purposes of this work) is read into Matlab prior to running the scripts.

In order for the thesis software to be used as a pathology tool, a user-friendly interface is required. Via this interface, the user should be able to select the slide of interest, and have the option of running all analysis steps on all areas of this slide in a single step. With the current algorithms and memory limitations of Matlab, it appears that processing a slide as a series of smaller images is a necessary restriction. A script (*processImage.m*) has been developed which selects areas of the overall slide image in successive 2000 x 2000 pixel blocks, and runs the alveoli and inflammation algorithms on each block in turn. Since only a limited number of areas of each slide were analyzed for the thesis project, this script was not used. Functionality similar to this, bundled with a user interface, is required. The script would need to be expanded to also run the fibrin and necrosis detection algorithms.

Due to the lengthy run times for processing images, in its current form the tool could not be interactive. The user would be required to select the slide(s) of interest and start the processing. An alert of some kind could be used to signal when the processing was complete. A user interface for easy selection and viewing of the highlighted images created as output by the software is also required. Statistics should also be consolidated from each of the individual analyzed image regions to produce a single report for each full slide.

One key area for usability improvement is in the excessive processing time required by the current Ojala-Chen algorithm, as implemented in the thesis software. The Ojala-Chen script is an intermediate step in locating areas of fibrin and necrosis in the images. When run on dedicated dual CPU nodes with 2.4 GHz Xeon processors and 2 GB of RAM, this one script required from a low of about 5 hours to a high of 122 hours elapsed time to complete analysis of a 2000 x 2000 pixel image. These long processing times occurred even after a number of coding efficiency improvements dramatically reduced the required run-time of this algorithm (run-time of the region merging step dropped by a factor of 60 as a result of coding improvements, for example). Of the three major steps of this algorithm (colour quantization, hierarchical splitting, agglomerative merging), it is the merging step which requires the majority of total processing time. Additional coding efficiency improvements may be possible, but one key way to cut run-time would be to further limit the size of the minimum area in the hierarchical splitting step. Currently, images can be divided into regions as small as 12 x 12 pixels. Increasing this minimum size would reduce the number of image regions created by hierarchical splitting which may have some impact on the processing time of this step. More significantly, a reduction in the number of image regions to evaluate would improve processing time of the agglomerative merging step. More investigation would be required to determine an appropriate minimum block size, such that processing time is reduced, but accuracy of the final regions is not overly sacrificed.

Another possibility worth exploring is the potential improvement in run-time of these algorithms if they were executed on massively parallel shared memory multi-processor machines, such as available on WestGrid (www.westgrid.ca).

6.4 Final Comments

Although it is possible to make some improvements through minor adjustments to the existing algorithms, and to potentially improve results by implementing alternate or additional image processing techniques, before any further work is undertaken two key areas must be addressed. First, a more thorough understanding of the problem domain is required. Analysis of the manual assessment results revealed a number of areas in which the software did not allow for the full range of variance in appearance of some of the disease indicators, or for the number of possible combinations of features within the same image region. The programs will never successfully detect features which they were not trained to find. Since a substantial amount of pathology knowledge is required in order to define the full set of possible cases which the software should identify, the ideal situation would have a pathologist working directly with the software developer to define a full and thorough set of training images, to analyze program results, and to provide continuous feedback for improvements. Secondly, a better way of evaluating the programs' results must be determined. Counting regions of a certain feature and basing the evaluation on sensitivity and false positive counts has a number of shortcomings, as discussed previously. A better way to measure the accuracy of assessment results when compared to an expert analysis is required. Comparison based on overall area occupied by disease indicators is one possibility, but this comparison alone also would not indicate the full accuracy or inaccuracy of the programs.

Automated analysis of histological images is a difficult problem. The huge variations in the biological structures make even defining a set of target characteristics challenging. The fact that expert histopathologic analyses themselves are subjective and do not follow an established set of steps means that there is not a generally accepted process that the software can emulate. Still, even though the thesis software is not yet in a state where it could be used as a tool for pathologists, it has potential. At the very least, the results of the software have shown that automated detection of disease indicators in lung tissue samples is a possibility.

LIST OF REFERENCES

[1] A. D. Ramsay, “Errors in histopathology reporting: detection and avoidance”, *Histopathology*, vol. 34, pp. 481-490, 1999.

[2] E. R. Farmer, R. Gonin, and M. P. Hanna, “Discordance in the Histopathlogic Diagnosis of Melanoma and Melanocytic Nevi Between Experts”, *Human Pathology*, vol. 27, no. 6, pp. 528-531, June 1996.

[3] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, Nov. 1973.

[4] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, 2nd ed., San Diego, CA: Academic Press, 1997.

[5] Z. Liu, H. L. Liew, J. G. Clement, and C. D. L. Thomas, “Bone Image Segmentation”, *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 5, pp. 565-573, May 1999.

[6] D. Griffin, “Economic Impact Associated with Respiratory Disease in Beef Cattle”, *Veterinary Clinics of North America: Food Animal Practice*, vol. 13, no. 3, pp. 367-377, Nov. 1997.

[7] G. A. Andrews, and G. A. Kennedy, “Respiratory Diagnostic Pathology”, *Veterinary Clinics of North America: Food Animal Practice*, vol. 13, no. 3, pp. 515-547, Nov. 1997.

[8] P. A. Bretscher, *The Principles of Immunology*, Notes for Microbiology 421.3 / 721.3, University of Saskatchewan, 2003.

[9] R. A. Harley, “Histochemical and Immunochemical Methods of Pulmonary Pathology”, *Histochemistry in Pathologic Diagnosis*, S. Spicer (Editor), New York, NY: Marcel Dekker Inc., pp. 265-331, 1987.

[10] C. Nagaishi, *Functional Anatomy and Histology of the Lung*, Baltimore, MD: University Park Press, 1972.

[11] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed., Pacific Grove, CA: PWS Publishing (Brooks/Cole Publishing Company), 1999.

[12] B. Jahne, *Practical Handbook on Image Processing for Scientific and Technical Applications*, 2nd ed., Boca Raton, FL: CRC Press, 2004.

[13] M. Tuceryan, and A. K. Jain, “Texture Analysis”, *The Handbook of Pattern Recognition and Computer Vision*, 2nd ed., C.H. Chen, L.F. Pau, and P.S.P. Wang (Editors), Singapore: World Scientific Publishing Co., pp. 207 – 248, 1998.

[14] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support Vector Machines for Texture Classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542-1550, Nov. 2002.

[15] D. F. Morrison, *Multivariate Statistical Methods*, 4th ed., Belmont, CA: Brooks / Cole Thomson Learning, 2005.

[16] S. K. Kachigan, *Statistical Analysis: An Interdisciplinary Introduction to the Univariate and Multivariate Methods*, New York, NY: Radius Press, 1986.

[17] T. Wurflinger, J. Stockhausen, D. Meyer-Ebrecht, and A. Bocking, "Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images", *Computerized Medical Imaging and Graphics*, vol. 28, pp. 87-98, 2004.

[18] T. Tanaka, Y. Ohno, and T. Oka, "Classification of Malignant Tumor using Image Processing", *World Congress on Medical Physics and Biomedical Engineering*, Aug., 2003.

[19] P. Pavlova, K. Cyrrilov, and I. N. Moumdjiev, "Application of HSV Colour System in Identification by Colour of Biological Objects on the Basis of Microscopic Images", *Computerized Medical Imaging and Graphics*, vol. 20, no. 5, pp. 357-364, 1996.

[20] K. Sutherland, and J. W. Ironside, "Automatic Texture Segmentaion using Morphological Features on Images of the Human Cerebellum", *Proceedings of the IEEE International Conference on Image Processing and Its Applications*, pp. 777 – 780, July 1995.

[21] O. Lezoray, and H. Cardot "Cooperation of Color Pixel Classification Schemes and Color Watershed: A Study for Microscopic Images", *IEEE Transactions on Image Processing*, vol. 11, no. 7, pp. 783-789, July 2002.

[22] K. Sutherland, D. Rutovitz, J. E. Bell, and J. W. Ironside, "Evaluation of a novel application of image analysis to spongiform change detection", *Proceedings of the IEEE International Conference on Image Processing*, pp. 378-381, Nov. 1994.

[23] R. S. Poulsen, and I. Pedron, "Region of Interest Finding in Reduced Resolution Colour Imagery – Application to Cancer Cell Detection", *Pattern Recognition*, vol. 28, no. 11, pp. 1645-1655, 1995.

[24] R. N. Sutton and E. L. Hall, "Texture Measures for Automatic Classification of Pulmonary Disease", *IEEE Transactions on Computers*, vol. C-21. no. 7, pp. 667-676, July 1972.

[25] T. Ojala and M. Pietikainen, "Unsupervised texture segmentation using feature distributions", *Pattern Recognition*, vol. 32, pp. 477-486, 1999.

[26] K. M. Chen and S. Y. Chen, "Color texture segmentation using feature distributions", *Pattern Recognition Letters*, vol. 23, pp. 755-771, 2002.

[27] R. R. Sokal and F. J. Rohlf, *Biometry*, 3rd ed., New York, NY: W. H. Freeman and Company, 1995.

[28] M. J. Swain and D. H. Ballard, "Color Indexing", *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, 1991.

[29] C. H. Lo and S. Y. Chen, "Image classification using cellular color decomposition", *Proceedings of the IPPR Conference on Computer Vision, Graphics and Image Processing*, Taiwan, vol. 2, pp. 681-688, 1999.

[30] K. W. Bowyer, "Validation of Medical Image Analysis Techniques", *Handbook of Medical Imaging: Volume 2 – Medical Image Processing and Analysis*, M. Sonka and J. M. Fitzpatrick (Editors), Bellingham, WA: SPIE Press, 2000.

[31] F. W. Samuelson and N. Petrick, "Comparing Image Detection Algorithms Using Resampling", *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano*, pp. 1312-1315, Apr. 2006.

APPENDIX A
SUPPLEMENTARY INFORMATION

Detailed results for all images are provided in electronic format accompanying this thesis. For each of the 50 slide regions analyzed, the five highlighted images produced by the software are included, along with both the Image Analysis Report and the Fibrin and Necrosis Report for these analyses. Highlighted images from the expert manual analysis are also included for each region. In addition, a Results Table document shows the detailed comparison of the automated and manual results for all image regions. (This is the same amount of detail for all image regions as is included in section 5.2 for three sample image regions.)

The files are organized as follows:

- The Results Table document is located in the root directory.
- There is one subdirectory for each of the 10 slides.
- Each slide directory contains 5 subdirectories, one for each image region.
- Each image region directory contains:
    - The original image
    - An "Automated Assessment" subdirectory which contains the highlighted images and reports produced by the software
    - A "Manual Assessment" subdirectory which contains the highlighted images produced by the expert manual analysis