Grading of Dysplasia in Barrett’s Oesophagus using Computer Vision
Ben Snape
BSc Computing
2008/2009

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Signature of student) ______________________________
Summary

Barrett’s oesophagus is a rapidly increasing phenomenon, with recent estimates of 1-5% of the population in Western countries suffering from the condition (Abrams, 2009). This rise has been strongly linked (Solaymani-Dodaran et al., 2004) with an increasing incidence of oesophageal adenocarcinoma. This risk was illustrated recently (Lepage et al., 2008) where it was found in an extensive retrospective 31-year study that the risk of oesophageal cancer for UK men and women has increased threefold since 1971. Furthermore, in recent months the National Institute for Clinical Excellence (NICE) have begun preparing guidelines on behalf of the Department of Health for the standardised treatment of Barrett’s.
Acknowledgements

Special thanks go to my inspirational supervisor, Dr. Derek Magee, Dr. Darren Treanor for his excellent advice and Dr. Andy Bulpitt for his input in weekly meetings also. I would also like to thank my assessor, Chris Needham, for his valuable comments and feedback.

A thank you must also go to my family for supporting me all these years, and my love to Carlene, who always has faith in me.
Contents

Introduction ................................................................................................................................. 1
  1.1 Aim ................................................................................................................................... 2
  1.2 Objectives .......................................................................................................................... 2
  1.3 Minimum Requirements ..................................................................................................... 2
  1.4 Project Management ......................................................................................................... 3

2. Medical Background ............................................................................................................... 4
  2.1 Barrett’s Oesophagus ......................................................................................................... 4
  2.2 The Categorisation of Barrett’s Oesophagus .................................................................... 6
  2.3 The risks that are involved ................................................................................................. 6
  2.4 Why the correct diagnosis is so important ........................................................................ 7
  2.5 Preparation of tissue for analysis ..................................................................................... 8
  2.6 Indicators for diagnosis ..................................................................................................... 9
    2.6.1 General Characteristics ............................................................................................... 9
    2.6.2 Low Grade Dysplasia ............................................................................................... 10
    2.6.3 High Grade Dysplasia ............................................................................................... 11
    2.6.4 Adenocarcinoma ....................................................................................................... 12
  2.7 Summary: Why Research is Important ............................................................................ 13

3. Computer Vision Background ............................................................................................... 14
  3.1 Image Pre-Processing ......................................................................................................... 14
3.1.1 Gaussian Smoothing ................................................................. 14
3.1.2 Colour Deconvolution ............................................................... 15

3.2 Related Computer Vision Concepts and Techniques ......................... 15
3.2.1 Connected components analysis .............................................. 15
3.2.2 Nuclei Detection ................................................................. 16
3.2.3 Delaunay Triangulation .......................................................... 16
3.2.4 Colour Histograms ............................................................... 17

3.3 The Difficulty of the Problem .......................................................... 17

4. Design and Implementation ................................................................. 19
4.1 Methodology .................................................................................. 19
4.1.2 The Demir and Yener (2005) Approach ....................................... 19
4.1.3 Chosen Methodology ............................................................... 20

4.1 Review of Technologies ................................................................. 20
4.2.1 MATLAB ............................................................................. 20
4.2.2 C/C++ ................................................................................ 21
4.2.3 Java .................................................................................. 21
4.2.4 OpenCV ............................................................................ 21

4.3 Image Pre-Processing ..................................................................... 21
4.4 Feature Extraction ......................................................................... 22
4.5 Implementation ............................................................................. 24
4.6 Learning ....................................................................................... 27
4.6.1 Colour Histogram Training ................................................... 27
4.6.2 Edge Length Histogram Training .......................................... 28

5. Evaluation of Results ...................................................................... 30
5.1 Nuclei Detection ......................................................................... 30
5.2 Classification Results .................................................................... 31
5.2.1 Colour-Based Classification .................................................. 31
5.2.2 Edge Length-Based Classification .......................................... 32
5.2.3 Combining the Two Classifiers ................................................................. 33

5.3 Future Improvements .................................................................................... 34

5.4 Conclusion ..................................................................................................... 34

References ........................................................................................................... 35
Introduction

Recent advances in cutting-edge technology has ushered in an era of Computer Vision based research applied to medical imaging, creating the new discipline of Medical Image Analysis (Ayache & Duncan, 1996). In recent years in particular, the state-of-the-art Aperio ScanScope CS scanner has made it possible to scan a $30\text{mm} \times 20\text{mm}$ piece of tissue and represent it digitally as a $120,000 \times 80,000$ pixel image, totalling $28\text{GB}$ (Wang et al., 2007). In tandem with the exciting opportunities that such advances bring, significant challenges such as coping with the large image sizes and detail, are inexorably introduced and must be overcome.

Barrett’s oesophagus is a condition whereby stomach acid forces its way into the lower oesophagus (Majumdar & Basson, 2006), gradually causing tissue abnormalities. Initially, this change appears to be a protective adaptation as the body replaces cells in the affected region with acid resistant ones (Orlando, 1999). However the price for protection is high because the tissue can then become pre-malignant (known as dysplasia), increasing the risk for the development of cancer (Bennett, 1996).

From a practical perspective, work in this field is primarily aimed at helping the pathologist, not replacing him (Van Vliet, 2003). However, as Van Vliet (2003) discusses, pathologists take many years of training to formulate an expert fuzzy logic methodology which at best is tedious, slow and subjective. Therefore, the clear goals in developing an automated system are objectivity, accuracy and efficiency.

Research into the area of Medical Image Analysis alone is vital for many reasons. In order to progress medically, great computational power is needed in combination with constant human
innovation. Increasing cancer survival rates relies heavily upon early detection and diagnosis of various cancers as well as their subtypes quickly and accurately (Waheed et al, 2007). Indeed, Waheed et al. state that expert knowledge of the behaviour of particular cancers and their corresponding treatment are key to clinical success. It has become clear that despite our apparent advances in combining the two over the past decade, there is still much work to do.

1.1 Aim

Barrett’s oesophagus is a potentially serious condition of the oesophagus caused by acid reflux, which if left untreated, can develop into cancer. Before this, however, it usually goes through a pre-invasive stage called dysplasia. The aim of this project will be to process virtual slides of suspect tissue, using computer vision, and conclude whether there is no dysplasia, low dysplasia, high or carcinoma present.

1.2 Objectives

1. Develop software to effectively classify images as either negative for dysplasia, low dysplasia or high dysplasia and carcinoma.

2. Develop an understanding of computer vision techniques with regard to tissue classification.

3. Evaluate the most effective methods for classifying an image using computer vision.

1.3 Minimum Requirements

1. Investigate methods for the classification of pathology slides using computer vision.

2. Develop software that classifies areas of tissue to a good standard based on the labelled training data provided.

3. Develop an evaluation protocol for the effectiveness of different tissue classification approaches.
1.4 Project Management

Appendix B shows the schedule that was made for the completion of the project. As it was not a typical project, specific methodologies were not followed. However, a method was still followed to some degree as outlined in section 4.
2. Medical Background

2.1 Barrett’s Oesophagus

Barrett’s oesophagus is a condition that affects the lower oesophagus, named after Dr. Norman Barrett who discovered the condition (Barrett, 1956). It is caused by acid and bile reflux originating from the stomach and gall bladder respectively, collectively known as duodeno-gastro-oesophageal reflux (DGOR) (Sharma & Sampliner, 2004), forcing their way through the stomach, past the lower oesophageal sphincter and into the lower parts of the oesophagus (Majumdar & Basson, 2006). This exposure of healthy oesophageal tissue to highly corrosive acid which it is unsuited to causes cell damage, and consistent exposure leads to significant tissue abnormalities (known as dysplasia) over a period of time. Thus, dysplasia can be defined as a localised, irregular growth of pre-cancerous, but generally treatable, cells (di Pietro et al., 2008).

More specifically, corrosive stomach acid comes into contact with oesophageal cells at the epithelial (surface) layer, which are of a non-keratinised stratified squamous (flattened) epithelium type (Majumdar & Basson, 2006). An example of this type of tissue can be found in Fig 2.1. The corrosive stomach acid forces the oesophageal cells to adapt to the new, acidic conditions, hence becoming more like stomach cells in a process known as metaplasia. These altered cells are no longer stratified squamous but nonciliated simple columnar, and have a completely different appearance. Fig 2.1 provides an example of these elongated cells. Spechler (2002, p 837) provides an excellent image that highlights both the physical and histopathological differences between a healthy oesophagus and one with Barrett’s.
This process of stomach-like, acid-tolerant columnar cells (Figure 1.b) replacing acid-damaged stratified squamous epithelium (Figure 1.a) appears to be a protective adaptation (Orlando, 1999). However the price for protection is high because this process greatly increases the risk of developing dysplasia (Majumdar & Basson, 2006), which in turn raises the likelihood for the development of oesophageal adenocarcinoma (Bennett, 1996).

It is important to note that these changed cells at the lower end of the oesophagus are not cancerous. However, these cells have an increased risk (compared to normal oesophagus cells) of turning cancerous in time. Recent research puts this increased risk at 30-40 fold (Majumdar & Basson, 2006).

Traditionally, Barrett’s oesophagus and its increasing frequency has been strongly linked (Corley et al., 2007; Eldelstein et al., 2007) with obesity due to increased abdominal pressure weakening lower oesophageal sphincter integrity. However, Kendall et al. (2008) dispute this simplistic view, who believe that other significant metabolic factors require further study, such as the hormone leptin. Despite this, more easily-implementable lifestyle modifications, such as reduced use or cessation of alcohol and tobacco has been recommended (Orlando, 1999) as well as reduced consumption of fresh fruit, seafood and milk (Fan et al., 2008) to limit the development of
oesophageal cancer. This certainly demonstrates that there is much still to be learnt regarding the true causes and development of Barrett’s oesophagus.

2.2 The Categorisation of Barrett’s Oesophagus

When diagnosing Barrett’s oesophagus, pathologists follow a six category scale (Treanor, 2008) which can be seen in Figure 5. However, for the purposes of this project, mainly on grounds of simplicity and increased accuracy, this six category scale will instead be simplified to a four category scale (Figure 5).

<table>
<thead>
<tr>
<th>1. Negative for dysplasia</th>
<th>1. Negative for dysplasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Indefinite (probably negative)</td>
<td>2. Low grade dysplasia</td>
</tr>
<tr>
<td>3. Indefinite (probably dysplastic)</td>
<td>3. High grade dysplasia</td>
</tr>
<tr>
<td>4. Low grade dysplasia</td>
<td>4. Intramucosal carcinoma</td>
</tr>
<tr>
<td>5. High grade dysplasia</td>
<td></td>
</tr>
<tr>
<td>6. Intramucosal carcinoma</td>
<td></td>
</tr>
</tbody>
</table>

(a) The Six Category Scale  (b) The Simplified Four Category Scale

Figure 2.2: The Traditional and Simplified Category Scales for Diagnosing Barrett’s Oesophagus

2.3 The risks that are involved

Barrett’s oesophagus has been subjected to considerable interest in recent years, mainly because there are a wide range of little-understood risks concerning the condition and its potential development into dysplasia or cancer. Firstly, it is still not properly understood why only some people who suffer from recurring acid reflux – known as gastro-oesophageal reflux disease (GORD) – develop the condition (Orlando, 1999). Much more importantly, neither is it known why only certain Barrett’s sufferers develop dysplasia or cancer (di Pietro et al., 2008). However, one certainty remains: early detection of the potential warning signs of low-grade dysplasia has been proved to significantly increase survival rates of a cancer with a typical prognosis of only 15% (Yousef et al., 2008).

Many studies (Miros et al., 1991; Orlando, 1999) have shown that most Barrett’s sufferers do not progress to cancer. Indeed, a new study showed that although patients with Barrett’s have increased mortality, only 10% die of oesophageal adenocarcinoma (Moayyedi et al., 2008). This contrasts greatly with the risk (in this study) of death from chest diseases (30%) and backs an earlier
study that claimed that most Barrett’s sufferers die of unrelated diseases and have a normal lifespan (Orlando, 1999). Moayyedi et al. (2008) furthermore make very clear the distinction between relative and absolute risk: namely that despite a twenty-fold increased risk of dying from adenocarcinoma as a Barrett’s sufferer, the actual number of deaths (25) were very low in their study (n=245).

It is important to note that ultimately, Barrett’s is a pre-malignant, but treatable condition (Orlando, 1999). However, the problem is that people are most often diagnosed with the condition after it has been allowed to progress to an advanced stage (such as high grade dysplasia). Often, this is because at advanced stages, patients display the more obvious symptoms such as weight loss and dysphagia (difficulty swallowing) (Rossi et al., 2006). When the condition is allowed to progress to such advanced stages – i.e. high-grade dysplasia or adenocarcinoma – the prognosis is very poor at below 15% (Yousef et al., 2008). However, it takes a number of years to progress from Barrett’s to potentially dysplasia or even adenocarcinoma. Currently no single genetic markers have been identified that can reliably predict whether a Barrett’s patient will develop cancer. Therefore regular and expensive patient observation and treatment is required, increasing significantly in terms of cost and regularity as the condition develops into dysplasia (Oberg et al., 2005). As a result, pathologists not only must deal with many patients whose condition is unlikely to deteriorate but also be constantly vigilant to the slightest sign of abnormalities: for example in a recent study it was found that the degree of high dysplasia present in a patient was irrelevant to the

Various studies over varying timescales and scope (Miros et al., 1991; Travis, Taylor & Misiewicz, 1998; Oberg et al., 2005; Rossi et al., 2006) broadly agree in their assessment of risk of Barrett’s sufferers developing low-grade dysplasia: approximately 10-30%.

2.4 Why the correct diagnosis is so important

It is important to understand that under- or over-diagnosing must tried to be avoided. Firstly, this is not only for financial and strategical reasons (Bennett, 1996), as costly and time-consuming treatment and observation is often radically stepped up or down between the various classifications of Barrett’s. Much more importantly, patient’s lives may be at risk, especially considering the recommended course of action in cases of high grade dysplasia being an oesophagectomy (removal of the oesophagus) which carries high mortality (7%) and morbidity rates (Orlando, 1999). However, in around a third of cases (Spechler, 2002; Sharma & Sampliner, 2004), by the time high grade dysplasia is diagnosed, cancer is also present.

Low grade dysplasia is seen remarkable only so far that it may develop into high grade dysplasia and possibly cancer. Indeed, the focus was much more often centred on high grade
dysplasia, such as by Buttar et al. (2001 cited in Srivastava et al., 2007, p.484), who reported that the risks of high grade dysplasia were the same regardless of extent. A later study (Dar et al., 2003 cited in Srivastava et al., 2007, p.484) disagreed with this conclusion, along with Srivastava et al.’s (2007) more recent investigation. When comparing the results of the two studies, Srivastava et al. (2007) found that the extent of low grade dysplasia forms a significant risk factor alone, leading to the development of oesophageal cancer. Consequently, those patients with extensive low grade dysplasia present are much more likely to develop cancer than was previously thought. While this project does not focus on extent, this could provide an interesting extension for further research.

2.5 Preparation of tissue for analysis

Once a patient is suspected of suffering from Barrett’s oesophagus, conclusive diagnosis is required. In order to achieve this diagnosis, human tissue is removed from the patient’s oesophagus with an endoscope by a doctor trained in its use. A camera is attached at the end of the endoscope along with surgical forceps to enable a suitable biopsy (tissue sample) to be removed if presence of Barrett’s is suspected. After the biopsy has been removed, it is sliced into cross-sections, then stained with haematoxylin and eosin (H&E) to stain the nuclei purple (haematoxylin) and the cytoplasm pink (eosin). Then it is digitally scanned by the Aperio ScanScope Scanner. For how a typical slide appears, please refer to figure 2.4 below.

Figure 2.4: How a typical slide looks from 1x magnification
Once the digital slide has been created, purpose-built software such as Aperio’s ImageScope mimics the capabilities of a traditional microscope, allowing tissue to be magnified up to the native resolution (usually 40X) and quick and easy panning across an image (Aperio, 2008).

2.6 Indicators for diagnosis

A fundamental part of the project was extensively researching and thus developing a solid understanding of the intricate differences between the various classifications of the condition. Examination of a wide range of literature, backed up by discussions with the pathologist, were combined in order to achieve the thorough understanding necessary to research possible computer vision methods required to solve the problem.

2.6.1 General Characteristics

A number of observable characteristics are present in what is known as simple Barrett’s – i.e. where there is no dysplasia present. These characteristics are important to identify because they change – often radically – at both the onset of dysplasia and its increasing severity.

Referring to figure 2.5, many features are clearly observable. Firstly, mucus-secreting goblet cells are plentiful. Secondly, a clear villiform architecture (finger-like projections) can be seen.
2.6.2 Low Grade Dysplasia

Barrett’s oesophagus with distinct low grade dysplasia begins to exhibit a change in architecture. Most notably, there is a significant reduction of mucus-secreting goblet cells (Sharma & Sampliner, 2006). Furthermore, a deterioration of the villiform architecture, leading to the formation of crypts is apparent (Falk & Goldblum, 2007). These crypts (figure 2.6) are remnants of the villiform architecture seen above in figure 2.5.
2.6.3 High Grade Dysplasia

High grade dysplasia continues the trend with a significant change in architecture and more pronounced nuclei clustering (figure 2.7). Additionally there is observable colour change due to inflammation.
2.6.4 Adenocarcinoma

The onset of adenocarcinoma, following on from high dysplasia, continues the trend of larger, more irregular and more crowded nuclei. Additionally, adenocarcinoma has the added characteristic of cribriform (perforated) tissue. Figure 2.8 below outlines these features.

The tissue will in places look strikingly similar to high grade dysplasia; indeed as adenocarcinoma follows on from high dysplasia, only small invasions of cancer may be visible. Despite this, in some cases the invasion of cancer is so extreme that to the untrained eye, the tissue may look like a region of low or no dysplasia.
2.7 Summary: Why Research is Important

A pathologist takes many years of training (Van Vliet, 2003) to learn a process that is highly subjective. There is proof (Brown et al., 2008) that the incidence of Barrett’s and adenocarcinoma of the oesophagus is rising, with up to 5% of the population in Western countries being affected (Abrams, 2009). Conio et al., (2001) suggest that there are five unidentified cases of Barrett’s oesophagus for every case that is diagnosed. Despite the advance of modern medicine, increasing cancer survival rates relies heavily upon early detection. So it is important to detect various cancers, as well as their subtypes, quickly and efficiently. Indeed, knowledge of the behaviour of particular cancers along with their treatment is key to clinical success (Waheed et al., 2007).
3. Computer Vision Background

3.1 Image Pre-Processing

Image pre-processing was the necessary first step required in order to prepare images for further analysis. A number of techniques are available.

3.1.1 Gaussian Smoothing

Gaussian smoothing is a technique that not only suppresses noise but also changes the intensity variation of the image; the latter of which suppresses or removes detailed features (Wink & Roerdink, 2004). Consequently, this may be an effective method of pre-processing the very large and detailed images.

The technique itself works by updating pixel values based on averaging the intensity values in the surrounding neighbourhood (Sonka et al., 2007). More specifically, it uses a kernel (small matrix) which contains values representing a discrete approximation of the Gaussian distribution in the form of a mask. This mask is moved across the image and each pixel is multiplied by the corresponding kernel value.
3.1.2 Colour Deconvolution

A method was recently developed in order to deconvolve (separate) the RGB colour of immunohistochemical staining, such as H&E (haematoxylin and eosin) allowing each stain to be identified and quantified separately (Ruifrok & Johnston, 2001). For example, this allows the nuclei-staining haematoxylin to be separated from the cytoplasm-staining eosin, leading to much more effective image analysis than previous methods which were based purely on HSV values which did not account for multiple stains contributing to a final colour (Ruifrok et al., 2003). Figure 3.1 clearly demonstrates the workings and accuracy of colour deconvolution in H&E images.

![Figure 3.1 – Haematoxylin and Eosin Separation](image)

(a) Deconvolved Nuclei (H)  (b) Deconvolved Cytoplasm (E)

3.2 Related Computer Vision Concepts and Techniques

Once sufficient image pre-processing has been achieved, relevant Computer Vision techniques must be implemented in order to classify test images. A variety of methods are available.

3.2.1 Connected components analysis

This technique involves the grouping together of adjacent, identical pixels in a binary image to form a single connected component (Wilson, 2006). From this, many results can be drawn such as the combined size of all connected components, as well as the size of the largest regions. These may highlight areas of densely clustered nuclei, which may indicate high dysplasia or cancer for example.
3.2.2 Nuclei Detection

Nuclei detection is a fundamentally important technique for many reasons. Not only does nuclei detection form the basis of many more advanced techniques (see 3.2.3), but it can act as an excellent classifier leading to accurate diagnoses. Therefore it is essential that the method is implemented to a high standard so that when more advanced techniques that rely upon nuclei detection are used, this accuracy is built upon and not hindered.

Although the method can be difficult to implement, there are a wide range of potential solutions available, ranging from the highly complex to less so. Fortunately, some excellent implementations are available (Swainston, 2008). It is hoped that available solutions will perform to an acceptable standard so that more difficult issues, should they arise, can be dealt with.

3.2.3 Delaunay Triangulation

Delaunay triangulation (Delaunay, 1934) is a mathematical technique, based on Voronoi diagram theory (Voronoi, 1908), which has been extensively used for many years to solve various scientific and engineering problems (Lee & Lam, 2008). Delaunay triangulation can be described as a triangulation for a given convex hull consisting of points in the Euclidean plane so that \( \{ p_1, \ldots, p_n \} \) and no two triangles intersect except share an edge, and the circumcircle of all triangles do not contain a point inside either (Scott et al., 2007).

Once Delaunay triangulation has been performed, calculation of edge lengths can be done using Pythagoras’ Theorem \( c = a^2 + b^2 \) where the two edges \( a \) and \( b \) between points are known. Figure 3.2 illustrates Delaunay triangulation and its relation with a Voronoi diagram and an example circumcircle.
The method has been employed to considerable success before when dealing specifically with digital pathology slides (Swainston, 2008), where it was used as the main classifier in achieving a respectable 67% average accuracy. Indeed, referring to clearly-identifiable characteristics (section 2.6): namely a deterioration of villiform architecture and increased nuclei crowding as Barrett’s with dysplasia progresses, should together result in observable disparity between edge length histograms of each class.

Implementing Delaunay triangulation sometimes results in anomalies in the form of long triangle edges at the periphery of a convex hull. Although solutions do exist, this will only be explored if it is deemed to be a significant stumbling block.

3.2.4 Colour Histograms

Colour, rather than intensity (represented in greyscale), is an important method in digital image processing. A particularly interesting approach by Sabo et al. (2006) took many factors of nuclei heterogeneity into account, most notably the optical density. It was believed that a similar method of measuring optical density, but of cytoplasm with the nuclei having been removed from the image, could be possible.

From observations of the tissue characteristics, such a method was seen as quite feasible. If it was possible to examine a wide range of images to create well-developed training sets, a good accuracy may be reached.

A number of potential problems exist however, mainly related to the tissue preparation stage. Over-staining and artefacts (such as tearing of the tissue) on the images were omnipresent issues. Inflammation of the tissue may also pose problems, leading to the solution confusing between the classes. This is compounded by the natural variation from person to person with regard to the absorption of H&E. Despite these worries, as the approach would involve the extensive use of histograms, it was believed that it could be very useful.

3.3 The Difficulty of the Problem

There exists a multitude of reasons why this particular problem is difficult. They range from the complex issues surrounding the biology of the condition, to the feasibility of what automated solutions will be able to provide. Additionally, the expectations of what an automated system could reasonably achieve when compared with other factors must be rigorously examined.
The project does not centre around a simple classification problem; deliberately contrived solutions will not be acceptable. Indeed, research in the past (Waheed et al., 2007) has focussed on very high achieved accuracy but in a non-automated and limited setting. Similarly, other research efforts have based systems of detecting two extremes of diagnosis, such as cancer against no cancer. Such systems have little potential for clinical use.

The classification problem itself is difficult for a number of reasons also. The six possible diagnoses for Barrett’s oesophagus, which although can be simplified to four due to the scope and level of this project, pales in comparison to projects mentioned above. In reality, even expert pathologists are prone to disagreeing when diagnosis Barrett’s due to its highly subjective nature (Demir & Yener, 2005; Spechler, 2002). Cell differences between each classification, and even intra-classification can be extremely subtle.

Bearing the last point in mind, inter-observer agreement of low dysplasia alone by expert pathologists has been reported as less than 50% (Spechler, 2002). Spechler makes clear that low dysplasia in Barrett’s oesophagus cannot be diagnosed reliably. As pathologists use a number of “hard” and “soft” cell identifiers in order to base their diagnoses upon, it can be difficult to know which features will be the most useful (van Vliet, 2003). Furthermore, many of these methods are simply too subtle to identify, or too complex to implement.
4. Design and Implementation

4.1 Methodology

Being a piece of scientific research rather than a traditional software engineering project has meant that an alternative frameworks had to be sought. Thus, methodologies for this line of work were researched in order to provide a clear structure of work to be carried out.

4.1.1 The Arif and Rajpoot (2007) Approach

Arif and Rajpoot (2007) outline their proposed algorithm (figure X), which was designed to meet their needs of detecting nuclei in an unsupervised learning approach. Being a research topic of similarity to the author’s it was felt that their method could be applied.

4.1.2 The Demir and Yener (2005) Approach

The approach consisting of three main steps (pre-processing, feature extraction and diagnosis) in order to classify cancer is a clear and well-structured method (Demir & Yener, 2005). They identify
the fact that the three steps heavily rely on each other, with the pre-processing stage perhaps being the most difficult and important.

4.1.3 Chosen Methodology

The author felt that both methodologies offered useful ideas, particularly as each were concerned with a very similar area of research. The former approach (Arif & Rajpoot, 2007), was similar to the latter (Demir & Yener, 2005), but it was felt that the more simple, yet instructive, method of Demir and Yener would be used (figure 4.2). The author identified particularly strongly with the sentiments regarding the vital importance of the image pre-processing.

![Diagram](image.png)

Figure 4.2: The Chosen Methodology

4.1 Review of Technologies

A number of potential technologies were available to implement the solution.

4.2.1 MATLAB

MATLAB is well known for providing a highly flexible, yet comprehensive environment with extensive documentation available. It is very popular in many areas of Computing and Engineering, such as digital image analysis. It offers excellent implementations of complex techniques (e.g. Delaunay triangulation) which can be employed quickly and easily with the in-built high-level programming language. Due to this, along with its gentle learning curve for even the novice programmer, it is often the technology of choice for research in this field. Its relatively friendly user interface at times hides the complexity of solutions possible in MATLAB.

However, there is sometimes a considerable trade-off between intuitiveness and ease of prototyping solutions with computational performance and efficiency; “for” loops are known to be especially inefficient. However, remedies exist in the ability to relatively easily embed other programming languages in MATLAB, most notably C/C++ code in the form of MEX files.
4.2.2 C/C++

C and C++ are very popular, and possibly the most well-known, programming languages. Consequently, they are also very popular in Computer Vision implementations. As the programmer has direct access to the underlying hardware, solutions are very fast and efficient. However, there is a much steeper learning curve associated with the language (and indeed any similar languages) when compared with MATLAB. In order to create solutions on a similar scale to those in MATLAB would require much greater technical expertise and a longer timeframe. Nevertheless, elements of C and C++ can be utilised in MATLAB to increase efficiency (see above).

4.2.3 Java

Java, in many respects, is similar to C and C++. It could be said that Java is a higher-level programming language than both, despite being derived from them. The strengths of Java lie in its extensive graphical library (known as Swing) – which can be used to create highly-functional graphical user interfaces (GUI’s) – as well as not being tied down by architecture with the inclusion of a Java Virtual Machine (JVM).

As is the case with C and C++, Java is fully-embeddable in MATLAB. This may be useful especially if a GUI needs to be created.

4.2.4 OpenCV

OpenCV is a powerful, yet not widely-used, image processing library for Computer Vision. It is more likely to be used in academia due to the research-oriented users who submit algorithms to the owner Intel. It is a very mature library, with most well-known algorithms being supported. It provides significant performance enhancements due to its low-level nature as well as having many available, flexible functions.

4.3 Image Pre-Processing

In order to prepare images for further analysis, they first had to be pre-processed (section 3.1). Gaussian smoothing was used first in order to reduce the very high level of detail of the image being examined. Figure 4.3 clearly demonstrates the reduction in excessive detail achieved when using Gaussian smoothing.
Figure 4.3: The Benefits of Using Gaussian Smoothing

Now, either further image pre-processing techniques can be used, or more advanced Computer Vision methods can be implemented directly. Bearing the former in mind, it is very desirable to perform colour deconvolution in order to separate the immunohistochemical stains present in the tissue, resulting in much more effective image analysis (section 3.1.2). Figure 4.4 compares using colour deconvolution to a simplistic greyscale conversion in MATLAB.

Figure 4.4: The Benefits of Using Colour Deconvolution

4.4 Feature Extraction

Identified features (section 2.6) were prioritised and pursued based on the background Computer Vision theory (section 3.2). Certain features, after discussion with the pathologist, were deemed to be
beyond the possible level and scope of the project. For example, the cribriform architecture present in regions of cancer – whilst observable – was deemed too difficult to identify automatically. Similarly, mucus-secreting goblet cells present predominantly in areas negative for dysplasia (but not in all cases), were perceived as “soft” indicators for diagnosis by the pathologist. By this it was meant that presence – or lack of – was not a reliable feature, but a supporting one. Whilst ideally extracting as many features as possible was desirable, goals had to be realistic; and thus only features of the highest integrity were pursued.

Features that were believed not only to be pervasive but subtly different throughout the various classifications (figure 2.2) were the ideal candidates. This meant that clearly distinguishable, omnipresent and extractable features, such as nuclei, were likely to be useful. Nuclei also exhibited small yet evident variations throughout the classifications: namely increased crowding correlating with the seriousness of the condition (figure 4.5).

![Increased Nuclei Crowding Correlating with Seriousness of Condition](image)

(a) No Dysplasia       (b) Low Dysplasia       (c) High Dysplasia       (d) Adenocarcinoma

Figure 4.5 – Showing the Relationship between Nuclei Crowding and Seriousness of Condition

One good feature had been identified but ideally, more were needed. It was decided that the second classifier should be colour-based, following the background research on other techniques (section 3.2.4). Consulting figure 4.5, a clear pattern can be observed in terms of subtle variations in colour between the classes. Typically, tissue would be stained very lightly with at times, considerable whitespace present. As dysplasia progressed, the cytoplasm became more pronounced as the nuclei became more irregular.
4.5 Implementation

With a clear view in mind regarding which features to implement, the project switched focus to utilising the researched techniques. A number of simplistic approaches were designed first in order to gain a better understanding of the issues.

Firstly, considerable experimentation with connected components ensued. Given an unclassified image, it was first pre-processed using Gaussian smoothing and colour deconvolution. Thresholding using Otsu’s method was consequently applied to the greyscale image to create a binary one. Some interesting results were observed (figure 4.6). However, the most significant thing observed was that the typical colour deconvolution vectors designed for H&E staining was doing a very poor job in dark areas – such as in high dysplasia and cancer.

![Figure 4.6: Showing Connected Components Analysis Results](image)

(a) High Dysplasia  (b) No Dysplasia

Connected components analysis, despite showing some promise was decided to be much too simplistic. Also, as the results were gained from the colour deconvolution not working properly, it did not seem the correct course of action to continue. Because of this observation, efforts became focussed on solving it.

Much time was spent making relatively little progress trying to sort this out. But the author felt that until the issue had been resolved, the implementation could not continue. This refers particularly to earlier comments made regarding creating a contrived solution; the desire was to have a single automated method.
The issue was largely resolved thanks to a tool gratefully provided by Derek Magee. It allowed a re-estimation of the colour vectors used in deconvolution so that regions of much deeper colour could be better pre-processed (figure 4.7).

Following this, the program output a series of vectors which were updated in the colour deconvolution code to include (figure 4.8).

```c
... } else if (!strcmp(myStain,"DRM")){
/* GL Haem matrix */
MODx[0]= 0.553434;
MODy[0]= 0.715757;
MODz[0]= 0.425913;
/* GL Eos matrix */
MODx[1]= 0.460384;
MODy[1]= 0.761963;
MODz[1]= 0.455476;
/* Zero matrix */
MODx[2]= 0;
MODy[2]= 0;
MODz[2]= 0;
...
```

The rest of the implementation swiftly picked up pace. The first classifier to be designed was based on colour. The code written to automate the process was relatively simple yet powerful. An
A unclassified image was loaded into a function which enacted a series of automated steps. First, the image was blurred and the stain deconvolved. Following this the nuclei were segmented using the optimal threshold gained through extensive testing. Care was taken in this respect not to under-segment the nuclei; slightly over-segmenting instead was the preferred choice. The segmented nuclei were then used to create a binary mask which was first complemented, then re-applied to the original image so that the nuclei appeared dark black. A histogram was created based on this original image with the binary mask applied, which when combined with the previous step meant that the simple removal of the black bins in the histogram effectively resulted in only the cytoplasm featuring. Additionally, any white space or background pixels were also caught and removed in this procedure. Figure 4.9 shows an example nuclei mask (the colour is shown deliberately; it would normally be binary) and the result of applying that mask to the original image.

![Nuclei Mask (in colour)](image1) ![Image After Applying the Mask](image2)

**Figure 4.9: Use of an Automated Nuclei Mask**

Next, the relatively simple implementation of Delaunay triangulation was needed. Because nuclei detection had been used to good effect, it was the simple matter of triangulating them and displaying the results.

In a process similar to that above, a histogram for an unclassified image was created, consisting of edge lengths obtained from Delaunay triangulation. An example can be seen in figure 4.10.
4.6 Learning

Learning was based on creating normalised histograms for a set of training images for each classification in the revised scale (figure 2.2). Therefore each Barrett's classification a training histogram consisting of ten images was created.

4.6.1 Colour Histogram Training

It was hoped that this method would be able to separate effectively the different classifications. Figure 4.11 shows the four training histograms created.

Figure 4.10: Delaunay Triangulation in a Low Dysplasia Image
4.6.2 Edge Length Histogram Training

A similar method was pursued for training using Delaunay triangulation. In order to ensure that vector dimensions agreed when concatenating the histograms for each class, the size of edge lengths displayed was constrained to a number that would not affect results. See figure 4.12 below.
Figure 4.6: Delaunay Triangulation Training Histograms

(a) No Dysplasia

(b) Low Dysplasia

(c) High Dysplasia

(d) Adenocarcinoma
5. Evaluation of Results

Significant testing and obtaining of results was undertaken in order to fully understand where the strengths and weaknesses of the system lay. A conscious effort (section 4.5) was made throughout the project to tackle the more challenging issues at hand, rather than over-simplifying the process.

5.1 Nuclei Detection

Nuclei detection was one area of the project that could be relied upon to a large degree because it was done with existing code. However, when errors did occur they were more often because of false positives rather than false negatives (figure X). But as there is no perfect method, the results were deemed adequate enough. In most cases, the accuracy of the nuclei detection was far more reliant upon effective image pre-processing.
Errors in the nuclei detection were mainly present in high dysplasia due to its very dark stain. The problem was that increasing seriousness of dysplasia in Barrett’s did get observably darker but once cancer set in this wasn’t always the case. However, it was hoped that linking in nuclei crowding through Delaunay triangulation with colour histograms would remedy this issue and the two techniques would complement each other.

5.2 Classification Results

Extensive testing was undertaken in order to identify areas of weakness in the system. Firstly, the features working separately were tested, before being joined together and evaluated also.

5.2.1 Colour-Based Classification

Classification results based on cytoplasm colour alone was successful in some respects. Figure 5.2 shows the actual percentage success rate when compared against the pathologist.
At first inspection, the results appear disappointing. This feeling is compounded by the poor average success rate. However, one area of relative success is the good classification of images with high dysplasia. Identifying where the classifier was confusing the correct diagnosis is the necessary next step (figure 5.3).

Consequently we can see that the problem lies with mainly with the confusion of no dysplasia images as cancer. Surprisingly the reverse does not appear to be true, with cancer images being classified correctly half the time. Furthermore, as high dysplasia and cancer diagnoses have the same clinical outcome, it could be argued it had a 100% success rate.

### 5.2.2 Edge Length-Based Classification

After being employed in a similar fashion previously (Swainston, 2008), the results obtained ought to be similar, around the 60-70% mark. This may not be the case, however, as this project uses a less simplified classification scale. Testing yielding the following success rates (figure 5.4):

<table>
<thead>
<tr>
<th></th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60%</td>
</tr>
<tr>
<td>2</td>
<td>60%</td>
</tr>
</tbody>
</table>
Table 5.4: Showing the Success of Edge Length-Based Classification

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>50%</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>55%</strong></td>
</tr>
</tbody>
</table>

The success rate is similar to the results obtained using the other classifier and below expectations. Figure 5.5 illustrates where the classifier was confused.

Figure 5.5: Confusion Matrix of Edge Length-Based Classification

5.2.3 Combining the Two Classifiers

It was hoped that combining the two classifiers would yield improved results. For each test image, two histograms were produced. A very simplistic method, due to a lack of time, was used. If the two classifiers disagreed, the diagnosis would fall back on the edge length-based one (figure 5.6).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60%</td>
</tr>
<tr>
<td>2</td>
<td>50%</td>
</tr>
<tr>
<td>3</td>
<td>70%</td>
</tr>
<tr>
<td>4</td>
<td>70%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>62.5%</strong></td>
</tr>
</tbody>
</table>

Figure 2.6: Showing the Success of Both Classifiers Combined

Overall, the slight increase in performance was pleasing, but would benefit from more extensive testing. With considerably more training images being made available, the accuracy would likely
increase. Attempts could be made at simplifying the classification scale further still, either by joining high and low dysplasia together, or grouping high dysplasia with cancer.

5.3 Future Improvements

From extended background reading and the author’s own experience with this project, a number of new and partially explored features were identified that could be extended in similar projects for the future.

- The most notable of these refers to one of the original possible extensions. It was proposed that some form of GUI front-end connected to an image database back-end could be a very promising idea. This tool could serve as a pathologist training application which makes diagnoses on an image, explains why it has come to the decision and displays relevant images from its database to support the conclusion.

- An area that was explored, but was ultimately seen as surplus to requirements was a MATLAB function that given a slide number (from the online server) and a choice of magnification would extract patches of a given size (recommended 500x500). Furthermore, the function determined the level of whitespace in each patch in real-time, thus deciding whether an appropriate level of foreground (i.e. tissue) pixels were present to justify writing the image to disk. The concept was based on previous code (Swainston, 2008) but was very heavily modified. The author believes that using this function could help pathologists identify Regions of Interest (ROI’s) (Wang et al., 2007) with the straightforward addition of other simple classifiers. This would mean that entire slides of tissue could be examined by the system, which may identify a few dozen regions of particular interest for closer inspection.

- Now that multiple classifiers have been implemented, with perhaps the addition of an extra one or two, a much more sophisticated machine learning techniques ought to be implemented to improve classification results.

5.4 Conclusion

The area of research that this project falls under is difficult, but hugely rewarding. It is truly a reflection of the advances of modern technology that human tissue can be represented digitally and to such a high level of detail.

There is much potential for further research in this area and time will tell how powerful solutions may become.
References


Ayache, N. & Duncan, J., 1996. From the Editors. Medical Image Analysis, Volume 1, Issue 1, Page iv.


35


NM Rajpoot, M Arif, AH Bhalerao,Unsupervised Learning of Shape Manifolds,in Proceedings British Machine Vision Conference (BMVC’2007), September 2007 - learning different shape types based on shape


From the very beginning of the year when I saw the potential project titles posted on the VLE, I knew that I wanted to pursue a project in this area. I distinctly remember feeling completely taken aback at the possibilities of Virtual Pathology combined with Computer Vision. It was completely unknown to me at that stage that such things were possible. So great was my enthusiasm, I even had to dissuade my flatmate Gavin Ackroyd from applying for the same project.

I felt, and still feel, extremely fortunate that I had such an enthusiastic and approachable supervisor in Dr. Derek Magee. For any shortcomings in my project, the blame must only lie on me, as Derek was a constant source of potential ideas and encouragement. I also feel extremely fortunate for benefitting from the expertise of Dr. Andy Bulpitt alongside Derek in my weekly meetings. Also, the presence in meetings, and anytime access to a consultant histopathologist in Dr. Darren Treanor was a particular highlight. He was also approachable at any time and always full of ideas. My only regret is that I did not feel I made the most out of my privileged situation. I hope that Derek and Darren especially take comfort in the fact that it was due to my independent nature that I did not seek help at times, most certainly not a lack of enthusiasm.

Undertaking and completing this project has, in my eyes, vastly improved the confidence in my own ability. I learnt that time management is critical with regard to projects of this size and I will certainly take away many lessons.

I was most pleased about the level of understanding I believe I achieved in this project. Thorough research in both areas of the project was difficult to wholly justify in the report. Additionally, the amount of time spent testing and refining code and methods simply cannot be expressed, with the exception that I could probably claim squatter’s rights in the ENIAC lab.
To students who are undertaking a similar project, I would advise you to learn from my mistakes. Implementation cannot be started early enough, but background research is extremely important too. Setting small weekly goals will allow you to remain on-track and not become disoriented by the larger problems, much like my colour deconvolution issue.