

**Towards Automated
Analysis for Prediction of
Myeloproliferative
Neoplasms using
Computer Vision**

Harry Smallbone

*This report is submitted as partial fulfilment
of the requirements for the Honours Programme of the
School of Computer Science and Software Engineering,
The University of Western Australia,
2017*

Abstract

Two life-threatening diseases, essential thrombocythosis and primary myelofibrosis, are currently diagnosed manually using histological images of cells called megakaryocytes. Differential diagnosis criteria for these diseases set by the World Health Organisation (2008) have widely reported problems with inter-observer reliability. This is hypothesised to be due to an absence of quantitative, repeatable criteria that can be systematically applied by pathologists. This problem could be alleviated by the development of an algorithm that can extract quantitative megakaryocyte characteristics from the same images used by pathologists, compare them to a previous set of ground truth data and present a diagnosis.

The key megakaryocyte segmentation and classification paper in this field used mathematical morphology to segment and classify individual megakaryocytes. They categorised individual megakaryocytes into disease states with a recognition rate of 91.7% and 86.2% for essential thrombocythosis and primary myelofibrosis respectively. However, their dataset was diagnosed using criteria from 2001 and their methods did not take into account megakaryocyte clustering, patient-wide features or disease progression, all key elements of the 2008 World Health Organisation criteria. There is room for improvement in this research area.

To address the research problem a multi-stage algorithm was developed. Machine learning is used for pixel-level segmentation of histological images. Next, a vector of numerical features is extracted to build a feature database for each disease state. Supervised learning classifiers were trained on this database to come up with the best classification methodology for use by pathologists.

The best classifier achieved 93% classification accuracy on a real-world image dataset. Each image was semi-automatically segmented and automatically classified. The primary contributions of this project are thus: a methodology for segmenting, analysing and classifying images of megakaryocytes, an investigation into megakaryocyte clustering and an analysis of the best criteria found for the diagnosis of essential thrombocythosis and primary myelofibrosis.

Keywords: Medical Image Analysis, Computer Vision, Automatic Classification
CR Categories: I.4, I.5, J.3

Acknowledgements

First and foremost I would like to thank my supervisors, Professor David Smith and Professor Bruce Gardiner. David's patience in explaining medical terms and ability to consistently point me in the right direction in spite of the many setbacks we faced over the course of this project was invaluable and greatly appreciated. Bruce never failed to alleviate stress with his measured approach to the project, entertaining stories and helpful advice on computer science, biology and thesis structure. I am extremely grateful for David and Bruce's meticulous feedback on my drafts despite time pressure and their diverse pieces of advice over the course of this year.

Secondly, this project would not have been possible without data and support from the Translational Cancer Pathology Laboratory of the University of Western Australia. Wendy, Kathy and Bob all put up with my requests for different microscopy and staining techniques and provided insight into the pathology and diagnosis process. I would particularly like to thank Bob for cutting and staining hundreds of bone marrow sections for me while investigating different microscopy techniques. Thank you to Kathy for providing the final dataset and checking my thesis draft.

I would also like to acknowledge the Centre for Microscopy, Characterisation and Analysis who provided access to different microscopes and investigated the possibilities for different data sources. Alysia and Paul's expert opinions and training were crucial at various stages of this thesis, and I am thankful for their time.

Finally, I would like to thank various friends and family who encouraged me, kept me sane and listened to my thoughts throughout the year. In particular, Ash for providing feedback and experiences from his own Honours thesis, and Saiuj, Mitch, Katie and Matt for helping me relax throughout an otherwise stressful year.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
2 Literature Review	7
2.1 Criteria Selection	8
2.1.1 Background	8
2.1.2 Problems Applying the WHO Criteria	10
2.1.3 Alternative Criteria	10
2.2 Image Acquisition	11
2.2.1 Optical Microscopy	11
2.2.2 Confocal Microscopy	12
2.3 Megakaryocyte Segmentation	13
2.4 Research Gap	14
3 Design and Implementation	16
3.1 Data Preparation	17
3.1.1 Input Data Selection	17
3.2 Segmentation	20
3.2.1 Convolutional Neural Network	21
3.2.2 Supervised Classifier Segmentation	23
3.3 Feature Extraction	25
3.3.1 Megakaryocyte Labelling	25
3.3.2 Extracted Features	26
3.3.3 Elliptic Fourier Feature	27

3.3.4	Cluster Detection	28
3.3.5	Classification and Criteria Analysis	29
3.3.6	Classification Algorithms	29
4	Evaluation	32
4.1	Dataset	32
4.1.1	Image Collection	32
4.1.2	Diagnosis Collection	33
4.1.3	Segmentation Training Set	34
4.1.4	Caveats	35
4.1.4.1	Image Defects	35
4.1.4.2	Training Set	36
4.1.4.3	Diagnoses	37
4.2	Segmentation	37
4.2.1	Training	37
4.2.2	Results	38
4.3	Feature Extraction	39
4.3.1	Extraction	39
4.4	Classification and Criteria Analysis	40
4.4.1	Machine Learning algorithms	40
4.4.2	Classifier Results	40
4.4.3	Clustering Results	42
4.5	Discussion	43
4.5.1	Segmentation	43
4.5.2	Classification	45
4.5.3	Clustering	49
5	Conclusions	50
5.1	Evaluation of Aims	50
5.1.1	Quantitative	50
5.1.2	Reliable	50

5.1.3	Robust	51
5.1.4	Practical	51
5.2	Future Work	51
5.2.1	Dataset Improvements	51
5.2.2	Better Diagnosis Grading	52
5.2.3	Additional Feature Extraction	53
5.2.4	Semi-automated Megakaryocyte Segmentation	53
5.3	Summary	53
	Bibliography	53
	A Glossary	58
	B Various Differential Diagnosis Criteria for Essential Thrombocytosis and Primary Myelofibrosis	61
	B.1 World Health Organisation Criteria	61
	B.2 Alternative Criteria	63
	C Statistical Measures	65
	C.1 Dice Similarity Coefficient	65
	C.2 Inter-observer Reliability	65
	C.3 F-Score	66
	C.4 Root Mean Squared Error	66
	C.5 Sensitivity and Specificity	66

List of Tables

4.1	The final image dataset, grouped by diagnosis	34
4.2	Segmentation results	38
4.3	Classification accuracy from each algorithm on individual megakaryocyte features	41
4.4	Classification accuracy from each algorithm on the bootstrapped megakaryocyte populations.	41
4.5	95% confidence intervals for the mean of each megakaryocyte feature (Elliptic Fourier coefficients omitted)	42
4.6	Average Silhouette Coefficients for each clustering algorithm . . .	43
4.7	Mean cluster statistics for each class using the affinity propagation clustering algorithm	43
B.1	The World Health Organisation diagnosis criteria for primary myelofibrosis, adapted from Swerdlow <i>et al.</i> [1]	62
B.2	The World Health Organisation diagnosis criteria for essential thrombocytosis, adapted from Swerdlow <i>et al.</i> [1]	63
B.3	Semi-quantitative diagnosis criteria for ET and PMF, adapted from Thiele <i>et al.</i> [5]	64

List of Figures

1.1	A simple depiction of megakaryocytes under different disease states	2
2.1	Sample images of megakaryocyte clustering in different disease states.	9
3.1	A flow diagram of the overall system structure.	17
3.2	An illustration of the Tissue Microarray approach	19
3.3	Sample CD61-stained bone marrow tissue	20
3.4	A diagram of the convolution neural network	23
3.5	A diagram of the Elliptic Fourier Descriptor process, adapted from Kuhl and Giardina [25].	28
4.1	A comparison of pre-fibrotic PMF tissue and fibrotic PMF tissue .	34
4.2	Sample image defects from the dataset	36
4.3	A sample segmentation of a well-formed megakaryocyte.	39
4.4	Sample megakaryocytes that were incorrectly segmented and necessitated manual segmentation	44
4.5	Comparison of the Elliptic Fourier feature at an individual and a population level	47
4.6	A graph of two important classification features with incorrect predictions shown in red.	48

CHAPTER 1

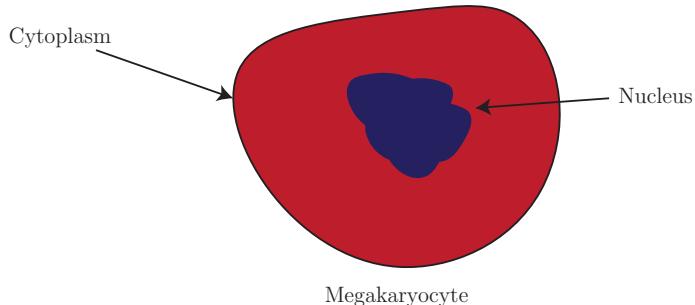
Introduction

Software is becoming increasingly important as a diagnostic tool for medical professionals, particularly in the area of medical image analysis. This thesis deals with analysing images related to two diseases, essential thrombocythosis (ET) and primary myelofibrosis (PMF). The diagnostic criteria for these diseases from the World Health Organisation (WHO) have well-documented issues when used in practice by pathologists. A research gap exists for creating an algorithm to build and apply quantitative criteria for differential diagnosis of ET and PMF. This gap is filled by contributing a semi-automated algorithm for analysing and classifying images of human bone marrow into different disease states, as well as an analysis of the various statistics used for classification.

Starting at a high level, the diagnosis of ET and PMF usually comes down to a systematic process of differentiating between the structure of specific cells in the bone marrow called megakaryocytes, known as differential diagnosis. The megakaryocytes of a patient are examined and categorised by pathologists using images of the bone marrow to come up with a diagnosis. This process, known as differential diagnosis, relies heavily on human judgement and experience to come up with consistent diagnoses.

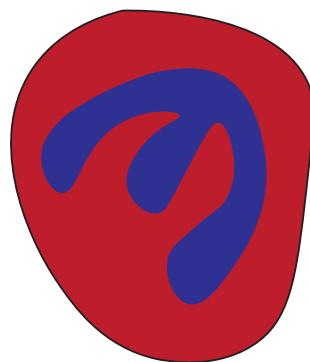
In normal bone marrow, megakaryocytes are roughly elliptical cells that tend to be quite spread apart from each other. In abnormal bone marrow, their shape, prevalence and size can vary dramatically. At a basic level the WHO criteria describe ET megakaryocytes as loosely clustered with ‘staghorn-like’ nuclei, and PMF megakaryocytes as densely clustered with ‘cloud-like’ nuclei and varying cytoplasm size [1]. A visual description of the differences can be seen in Figure 1.1.

Normal megakaryocyte
Infrequent megakaryocyte clustering, roughly elliptical nuclei



Essential Thrombocythaemia (ET)

Loosely clustered megakaryocytes
'Staghorn-like' nuclei



Primary Myelofibrosis (PMF)

Densely clustered megakaryocytes
'Cloud-like' nuclei

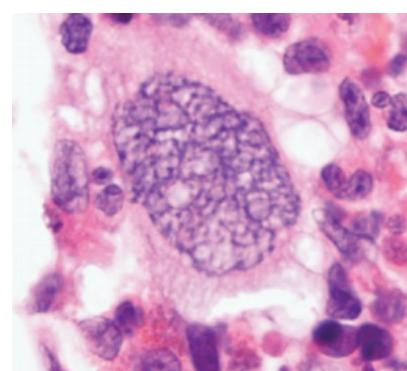
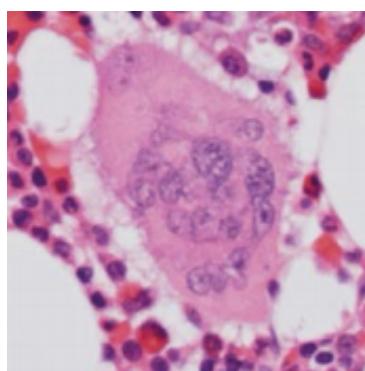
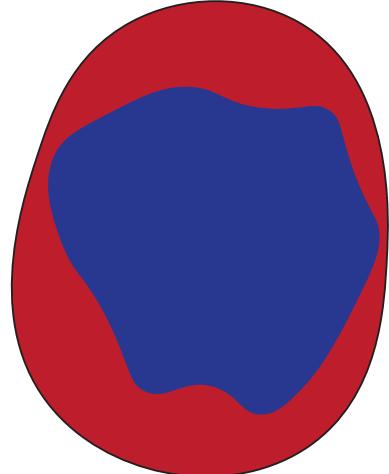


Figure 1.1: A simple depiction of megakaryocytes under different disease states. The red and purple colouring comes from chemical staining of the bone marrow. The images are adapted from Wilkins *et al.* [2, Figure 1].

Pathologists have had issues translating the WHO's qualitative criteria into reliable classification of ET and PMF, particularly in pre-fibrotic or early-stage PMF for reasons outlined later. This is addressed by providing quantitative descriptions of the features used for classification. The research area contains medical terminology which may be unfamiliar: these terms are explained inline or in the Glossary (Appendix A).

From a medical perspective, ET and PMF are categorised as myeloproliferative neoplasms (MPNs), which are rare, life-threatening diseases. They have a global incidence (probability of new MPN occurrence) of about 2.58 per 100,000 people per year and a reported prevalence (proportion of people affected by MPNs) of 93 per 100,000 in North America [3]. This means that in North America alone up to 330,000 people are currently affected by MPNs. There are four main MPNs that are most commonly diagnosed: polycythemia vera (PV), chronic myelogenous leukemia (CML), ET and PMF.

Both ET and PMF affect the production and morphology of megakaryocytes, which are cells in the bone marrow. This is life-threatening because megakaryocytes are responsible for creating platelets for circulation in human blood. Platelets circulate within the bloodstream and attach themselves to the walls of damaged blood vessels, where they staunch blood loss by activating fibrin clot formation, which is the first step towards stabilising and healing the damaged tissues. Abnormally high platelet counts can be life-threatening, as increased platelet concentrations can lead to spontaneous blood clots which may block blood flow to important parts of the body such as the heart or brain (this is particularly prominent in ET and early PMF). Decreased platelet counts causes reduced blood clotting around damaged vessels, resulting in excessive bleeding, especially in PMF.

PMF differs from ET in that it results in fibrosis (excess fibrous tissue formation) that increases in severity over time. The extra fibrous tissue interferes with tissue that would otherwise form new blood cells, impairing formation of megakaryocytes (and thus platelets) and all other red and white cells. This gives rise to the broad categorisation of PMF into 'early' or pre-fibrotic PMF which has similar symptoms to ET and 'late' or post-fibrotic PMF which is easier to diagnose due to widespread fibrosis and changes to blood cell production.

In 2008 the WHO defined semi-quantitative criteria to help pathologists distinguish between the various MPNs [1]. Clear diagnosis criteria exist for CML and PV based on blood tests due to their specific impacts on production of eosinophils and haemoglobin respectively. In contrast, pre-fibrotic PMF and ET may have similar blood profiles and so must be differentiated primarily by visually analysing bone marrow biopsies, looking for changes in megakaryocytes and mar-

row hyperplasia. The WHO recommends analysing the clustering, morphology and frequency of megakaryocytes to reach a conclusive diagnosis [1].

It is important to distinguish between the two diseases for prognostic and treatment reasons [4]. ET has a greater incidence compared to PMF of thrombocytosis (excess platelet production) and associated increased likelihood of blood clots, particularly as PMF develops and becomes fibrotic, but PMF has a significantly higher rate of leukemic transformation which contributes to increased likelihood of deaths following diagnosis of PMF [4]. The differential diagnosis of ET and early pre-fibrotic PMF is key to determining and managing potential risk factors and has a substantial significant impact on prognosis for patients [4]. The mean survival rate post diagnosis is 22 years with ET compared to 14 years with PMF according to one study [5].

The WHO criteria are reproduced in full in Appendix B and are detailed further along with some alternatives from the literature in Section 2.1. Briefly, the WHO describes ET patients as having loosely clustered large, mature megakaryocytes with hyperlobulated or staghorn-like nuclei, and PMF patients as having both densely and loosely clustered megakaryocytes with bulbous or cloud-like nuclei in a variety of sizes [6]. No definition of a cluster or cluster tightness is given by the WHO [1], but pathologists typically consider a megakaryocyte cluster to be a number of megakaryocytes in close proximity [2].

Despite the existence of the WHO criteria there has been controversy around how to effectively and practically apply them and so differentially diagnose of ET and pre-fibrotic PMF [6]. Inter-observer reliability for diagnosis, a measure of how consistent a diagnosis is across multiple pathologists, can range from 70% to as low as 35% [6]. It is hypothesised that this is due to a lack of quantitative criteria to aid in diagnosis: the WHO criteria [1] are qualitative when it comes to describing megakaryocyte morphology and do not provide any examples or weighting attached combinations of criteria in the key areas of megakaryocyte morphology and megakaryocyte clustering.

In this project, an algorithm is constructed for the automated segmentation and analysis of megakaryocytes in 2D bone marrow images. This algorithm is able to generate statistics to assist in the differential diagnosis of normal, ET and PMF bone marrow. To be practical, the implementation of the algorithm must meet the following criteria:

- *Quantitative*: as far as possible the software should report quantitative features of the sample megakaryocyte populations. Whereas a pathologist is only able to identify clusters of megakaryocyte semi-quantitatively (for example, describing a cluster as ‘loose’ or ‘tight’ and counting the number

of megakaryocytes), the algorithm should consider information on every megakaryocyte in a collection of images related to a patient and relate it to disease classification.

- *Reliable*: the software should be able to consistently identify and classify megakaryocytes populations with the highest accuracy possible.
- *Robust*: the images used in this project are preprocessed with a chemical stain to colour the megakaryocytes differently in the final image. This staining process is not exact and may result in different intensities and colours depending on how each individual tissue biopsy was processed. The software should be robust and cope with usual histological variations that occur during tissue preparation.
- *Practical*: there are a wide variety of existing techniques to analyse the bone marrow for information about its megakaryocytes. The software should follow the guidelines outlined in the WHO criteria and accept images prepared using standard pathology processes in order for it to be practical and useful to pathologists.

The algorithm was developed using bone marrow images sourced from the Translational Cancer Pathology Laboratory at the University of Western Australia. This work is presented in three chapters:

1. *Literature review*

A review was undertaken of the state of ET and PMF diagnostic processes and best practice in the areas of: bone marrow image acquisition, segmentation and analysis. By evaluating the various technologies in the context of megakaryocyte analysis it was determined that machine learning segmentation of 2D bone marrow images presents the best performance to flexibility tradeoff. The bone marrow images were chemically stained using CD61, an antibody that colours megakaryocyte cytoplasm in red which is useful for providing the segmentation algorithm with high contrast images.

2. *Design and implementation*

In this chapter the algorithm used to approach the research problem is described in three stages:

- Segmentation
- Feature extraction
- Criteria analysis

Firstly, the selection of the chosen image dataset and its corresponding diagnoses is explained.

Next, the chosen image segmentation methodology is described. The dataset images are segmented into labelled pixels in order to later extract regions of megakaryocyte cytoplasm and nuclei. A convolutional neural network segmentation algorithm was used to find regions of interest and a secondary machine learning segmentation algorithm fine-tuned the regions into high quality labelled megakaryocytes.

Regions of pixel labels are then grouped into megakaryocyte cytoplasm objects and megakaryocyte nuclei objects for feature extraction. As pathologists are still unclear on the relative importance of each WHO criterion, as detailed in the Literature Review (Chapter 2), an unweighted feature vector is constructed per megakaryocyte containing information such as area, perimeter and shape. The resultant feature vector is stored with the MPN diagnosis from the dataset. These per-megakaryocyte vectors are also aggregated to obtain a patient feature vector for patient-level classification. Megakaryocyte cluster information is analysed separately using a variety of clustering algorithms.

Finally, the criteria analysis stage has a supervised classifier learning component in order to find the best differential diagnosis criteria. The classifiers are tested and trained on both aggregated patient information and individual megakaryocyte features to find the most important features for classification. The clustering algorithms are evaluated and compared on the data but were not used for classification due to insufficient sample size.

3. *Evaluation*

A testing methodology is defined in this chapter for evaluating the performance of each stage of the algorithm under the input dataset. This involves building and testing a custom dataset for training the segmentation algorithms and the supervised learning classifiers.

The performance and other characteristics of the segmentation algorithms and supervised learning classifiers is described. The advantages and limitations of the approach are discussed and compared to the current state of the art WHO criteria and its various extensions in the literature. Finally, different ways to improve the methodology and extend it with further studies are explained.

CHAPTER 2

Literature Review

Differential diagnosis of essential thrombocythosis (ET) and primary myelofibrosis (PMF) is a complex and multi-level problem spanning issues of image acquisition, megakaryocyte feature extraction, data analysis and its application to the broader practice of pathology. In order to focus the investigation into the literature the research problem can be broken down into three interrelated areas of active research:

- *Criteria selection*: the WHO criteria for determining ET and PMF are qualitative and dependent on individual pathologist expertise for diagnosis [1]. Prior research has investigated the effects of various additional criteria on diagnosis consensus with mixed results.
- *Image acquisition*: the megakaryocyte features that can be extracted are highly dependent on the input images. Various techniques exist that have tradeoffs between image resolution, amount of information from the sample, method of preserving the source biological material and availability of training data to develop automated methods.
- *Megakaryocyte segmentation*: most imaging techniques also image cells other than megakaryocytes, thus requiring a segmentation process to label megakaryocyte cytoplasm and megakaryocyte nucleus pixels differently to the background. Since bone marrow images are noisy and differ greatly between patients, additional biological information is used to enhance the accuracy of these techniques. Previous work in this area has focused on both machine learning and traditional computer vision techniques.

The main megakaryocyte features previously examined in the literature are:

1. *Morphology*: the shape and size of the average megakaryocyte changes depending on whether the patient has normal, ET, pre-fibrotic PMF or late-stage fibrotic PMF. For examples, megakaryocytes with staghorn-like or cloud-like nuclei, and abnormal nucleus to cell area ratio.

2. *Clustering*: as a general rule MPNs increase the frequency of megakaryocyte in the marrow, and PMF causes many megakaryocyte clusters throughout the bone marrow. Imaging techniques that preserve positional information and can survey large areas of marrow provide more features for differential diagnosis.

The previous work in this area is examined to choose a combination of imaging technique, segmentation algorithm and diagnostic criteria to maximise the project's usefulness to pathologists in this field and its impact on diagnostic criteria.

2.1 Criteria Selection

2.1.1 Background

The full 2008 WHO criteria [1] for diagnosing ET and PMF are reproduced in Appendix B.

As a high level overview, the WHO criteria state that megakaryocyte morphology and atypia (structural cell abnormalities) are the primary distinguishing criteria between ET and PMF, specifically how clustered the megakaryocytes are, their shape, and their maturation stage [7, Table 5].

The full explanation is slightly more complicated. There are three major criteria defined by the WHO for PMF which must all be met for a true positive diagnosis: presence of megakaryocyte proliferation and atypia, fibrosis for late-stage PMF, the exclusion of other MPNs and either demonstration of a clonal marker (see Appendix A) such as the *JAK2* mutation or no evidence that the bone marrow fibrosis could be a secondary symptom of another disease [1]. Essentially the WHO require some evidence of atypical megakaryocytes and the exclusion of other diseases that could explain the symptoms. The presence of excess fibrotic tissue can positively identify PMF but this is not present in its early pre-fibrotic stages [1]. Thus in order to diagnose ET in the case of possible pre-fibrotic PMF the main comparison criteria is megakaryocyte formation and morphology [1].

There are no correlation coefficients, conditional probabilities or sample sizes for the current WHO morphological criteria. Furthermore there is no definition in the WHO criteria of what constitutes a megakaryocyte cluster beyond the qualitative description of 'dense' or 'loose' clustering [1]. There are no values for cluster incidence or prevalence in each disease state [1]. Some examples of

megakaryocyte clusters can be seen in Figure 2.1 but in practice, pathologists associate megakaryocytes into clusters based on past experience.

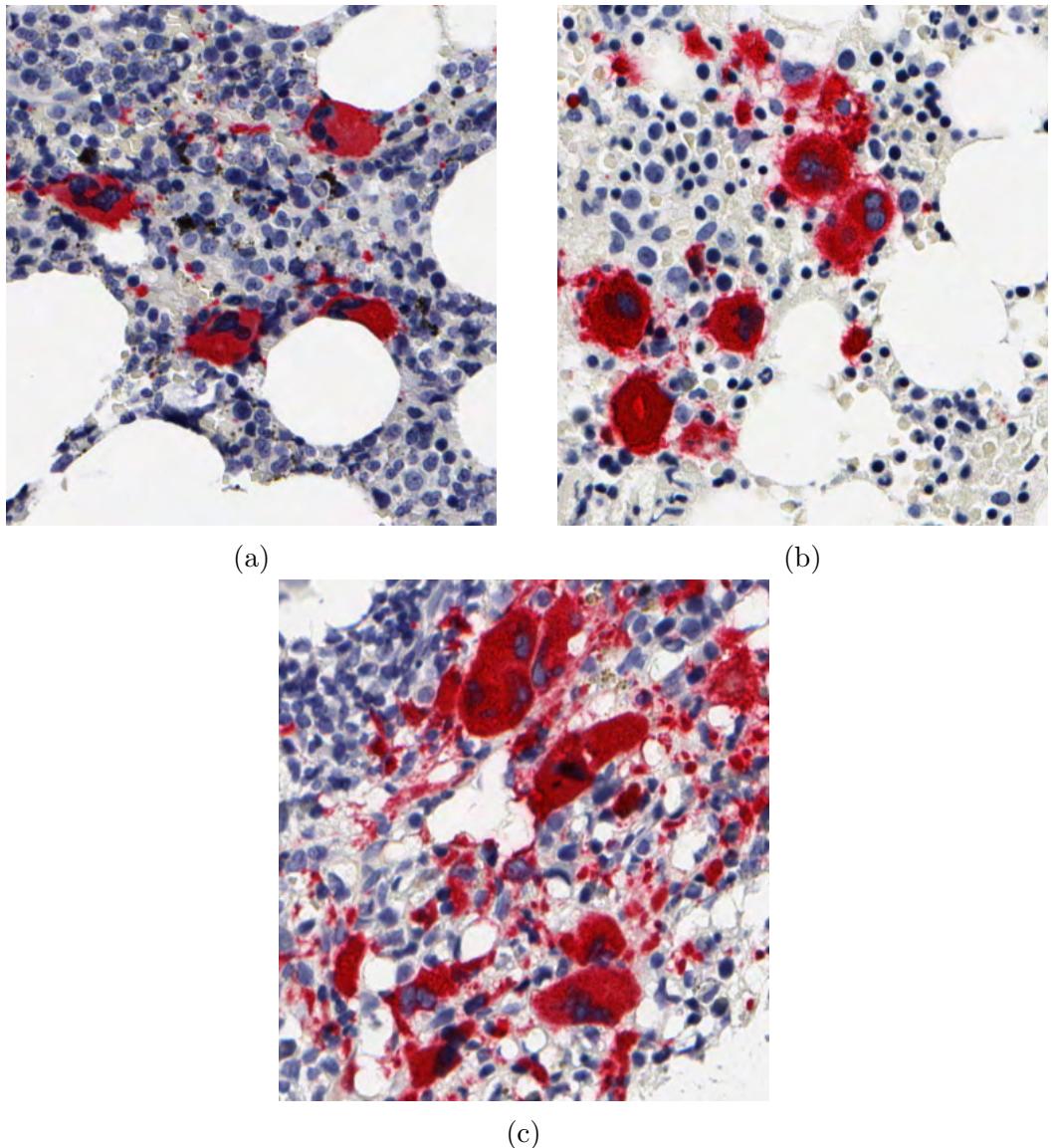


Figure 2.1: Sample bone marrow images showing the wide variety of clustering in different disease states.

(a) Some loosely clustered normal megakaryocytes. (b) Two sets of densely clustered megakaryocytes in essential thrombocythemia. (c) A number of megakaryocytes in close proximity in primary myelofibrosis.

2.1.2 Problems Applying the WHO Criteria

The literature shows some controversy around these criteria due to variable inter-observer reliability (see Appendix C) among practitioners using the criteria for diagnosis. In one study of 297 patients, only 78% (measured using Cohen's κ) of bone marrow image diagnoses using similar standardised criteria could be agreed upon by a group of pathologists [5], while another found only 53% agreement (using log-linear modelling of pairwise agreement) when applying the WHO criteria [2]. Barbui *et al.* [6] reviewed multiple studies of inter-observer reliability and found inter-observer reliability ranging from 35% to 63% to at most 70%.

In general, high inter-observer reliability and thus more accurate diagnoses are obtained by pretesting observers to ensure that inter-observer variability is minimised, and training them to ensure that the WHO criteria is applied as strictly as possible to the patients [6]. The primary problem seems to be a lack of concrete reproducible criteria that can be used to make a diagnosis, as well as the fact that it is difficult to follow this strict methodology in common practice due to the number of exclusion criteria that must be investigated [6].

2.1.3 Alternative Criteria

Some research has been conducted into developing quantitative criteria for the two diseases by studying how successful pathologists are in identifying diseases when given a list of features to look for. Some of these criteria are reproduced in Appendix B.

Thiele *et al.* [5] developed a list of semi-quantitative criteria for diagnosis, such as megakaryocyte cluster formation, size and presence of abnormal megakaryocytes, and gave them to pathologists along with relative probabilities of appearing in ET and PMF patients. This fully blinded survey resulted in a diagnosis consensus rate of 73% across 295 patients. This study illustrates the problems with current diagnostic processes: pathologists have to apply semi-quantitative or qualitative criteria to classify the disease, and the results are difficult to evaluate due to only testing inter-observer reliability instead of the proportion of correct diagnoses because the correct diagnosis is not always known without long-term followup.

Wilkins *et al.* [2] studied interobserver reliability using a set of qualitative criteria such as the presence of staghorn-like or cloud-like megakaryocytes assessed using a three-point scale and a calculated 'strength of association' for each criterion. Of the criteria, detecting bare megakaryocyte nuclei and the number

and size of megakaryocyte clusters was most reliable for diagnosis [2]. However, this study was heavily criticised by Kvasnicka and Thiele [8] for not measuring intraobserver reliability, having low levels on agreement on basic features like erythropoiesis (red blood cell production), taking smaller (0.5cm) than recommended (1.5cm) biopsies leading to less opportunity to observe clustering and for failing to standardise megakaryocyte features throughout the paper. Kvasnicka and Thiele [8] also discuss followup studies which showed that the original study group had a high proportion (~60%) of late-stage myelofibrosis patients which may have skewed the results.

The state of investigations into alternative criteria is summarised by Barbui *et al.* [6]. Problems that pathologists have experienced when applying the criteria, confusion around what constitutes pre-fibrotic PMF, and a lack of quantitative histologic features ordered by importance have all contributed to uncertainty about the differential diagnosis of ET and PMF [6].

2.2 Image Acquisition

In theory, many different imaging techniques such as X-ray microtomography, magnetic resonance imaging, optical projection tomography, optical imaging and confocal microscopy exist for image acquisition and analysis of bone marrow. In practice, imaging techniques considered here are restricted according to the following criteria: high cellular resolution (1-10 microns), ability to stain the cytoplasm/nuclei of megakaryocytes, as little tissue morphological alteration as possible during preparation and the ability to reconstruct 3D images. Techniques which are likely to enter clinical practice from research or that are already in use by practitioners to improve the practicality of any result are preferred. Of current techniques the two practical methods most suitable according to these criteria are optical imaging of 2D tissue sections and confocal microscopy [9] with the traditional 2D tissue sectioning approach being the most practical.

2.2.1 Optical Microscopy

Optical microscopy is the most commonly used method for acquiring images for use in differential diagnosis of ET and PMF. Travlos [10] provides an overview of this technique as accepted in modern practice: bone marrow biopsies are decalcified to remove opaque bone minerals, fixed into place on slides and imaged using an optical microscope. The WHO criteria were created using hematoxylin and eosin (H&E) stained optical sections as the basis for pathologists since they

preserve position and morphology characteristics as well as clustering by imaging large areas of the bone marrow.

H&E staining is the most common since it stains cell nuclei and cytoplasm differently allowing pathologists to compare megakaryocytes to the other cells such as eosinophils and leukocytes. A megakaryocyte specific stain can also be used to highlight megakaryocytes in a certain colour by utilising light-emitting fluorophores attached to antibodies which bind to a megakaryocyte-specific protein such as CD61. Megakaryocyte specific stains can make it easier to identify and analyse megakaryocytes for both humans and software.

Travlos [10] and Reagan *et al.* [11] list some disadvantages of this process, such as the need for precise decalcification and fixation protocols to avoid damaging cellular structure and inhibiting cellular staining [10]. Some artefacts can occur during the sectioning process, although great care is taken to try and avoid them. The fixative used to preserve the sample must be chosen carefully since widely used fixatives such as formaldehyde can shrink the sample by up to 4% or more [9] - this is acceptable as long as it shrinks the whole sample uniformly. Acid treatment is required to remove optically opaque minerals from the bone but damages protein epitopes in the process, which can weaken the effectiveness of antibodies depending on the extent of epitope damage. In some cases, the sectioning and staining process may have to be repeated multiple times to test which fixative and acid combination yields the best results [12]. Bone marrow can also be subjected to optical clearing agents which make opaque minerals transparent to avoid the tissue altering acidification process, but they are not considered here since it was not possible to obtain images with these relatively novel techniques.

Optical imaging is also strictly two dimensional which restricts the set of features visible by software such as three dimensional contours and clustering of megakaryocytes which cannot be extracted without taking multiple slides and manually registering them into a 3D volume. This also increases confusion about megakaryocyte features that can be extracted. For example, there is no definite way to know if a particular megakaryocyte is hyperlobulated or has simply been cut and imaged in such a way that it appears hyperlobulated.

2.2.2 Confocal Microscopy

Confocal laser scanning microscopy or confocal microscopy is an alternative imaging technique which allows for three dimensional optical sections to be imaged by capturing light emitted at a specific sample depth. This technique was recently applied to megakaryocyte imaging by Takaku *et al.* [13]. Due to the relative nov-

elty of the imaging process, no previous datasets of MPNs imaged with confocal microscopy exist and consequently there are no studies into its effectiveness for differential diagnosis of ET and PMF.

Confocal microscopy results in a stack of well-aligned 3D images, allowing it to capture three-dimensional features. By using protein-specific staining antibodies, megakaryocytes and their nuclei can be much more precisely and automatically separated from the background of the images. More of the imaging process can be controlled in software, such as the resolution in all dimensions and volumes of interest to capture different parts of the bone marrow. This means that morphology and clustering features can all be easily detected in three dimensions.

The principal limitation of this technique is the maximum thickness of the sample that emitted light can pass through. Takaku *et al.* [14] achieved sample thicknesses of 150 microns. While it is possible to image to this depth from both sides of the sample, this limitation restricts the potential number of visible megakaryocyte clusters and individual megakaryocytes which can be up to 100 microns thick. Another limitation is the inability to detect other biological material in the section which has not been tagged by antibodies even though it may be useful for diagnosis. Some examples of this are detecting the edge of the bone, and measuring the overall cellularity of the tissue, which have been used in some classification criteria [7]. Confocal microscopy is generally limited to a maximum of four different fluorescent frequencies per sample. This is because the fluorescent probes emit at a range of input and output frequencies. Widely used probes can have overlapping frequency ranges and subsequently be difficult to distinguish in the processing stage, commonly referred to as spectral bleed through. As a result, although the resultant image colours can be specifically associated with cell labels like megakaryocytes or granulocytes, only a maximum of four such colours are recommended with most available confocal microscopes. The marked cells and fluorophores used must be chosen with care to maximise potential sources of diagnostic criteria while also labelling megakaryocytes nuclei and cytoplasm separately.

2.3 Megakaryocyte Segmentation

The literature to date has focused on segmenting megakaryocytes from optical images of H&E stained bone marrow as outlined in Sections 2.1 and 2.2. For optimal feature analysis the segmentation algorithm must identify the exact pixels that belong to a megakaryocyte nucleus and belong to the surrounding cytoplasm, distinguishing them from the background pixels from other cells. This is made more difficult by varied megakaryocyte shapes and sizes, closely clustered

megakaryocyte, variations in staining colour and intensity and overlapping cells which can appear similar to nuclei.

Classical image segmentation techniques have already been applied to the problem of segmenting and classifying megakaryocytes. Ballarò *et al.* [15] obtained high classification rates of individual normal, ET and PMF megakaryocytes using a hybrid image segmentation algorithm combining wavelet transforms with standard image morphology techniques. They validated the algorithm using a database of 297 bone marrow microphotographs of 763x577 pixels taken at 400x magnification Ballarò *et al.* [15]. The database was specifically filtered for separated, mature, representative megakaryocytes for each disease state to avoid megakaryocytes with distortion or overlaps that could adversely impact its sensitivity [15], [16]. The final algorithm had a sensitivity of 98.4% for classifying normal versus pathological megakaryocytes and sensitivities of 91.7% and 86.2% for further classifying pathological megakaryocytes into ET and PMF respectively. Due to the nature of their dataset, clustering could not be analysed at all and no population-based analysis occurred.

Some machine learning techniques have recently been applied to the segmentation problem in isolation. Zarella *et al.* [17] trained support vector machine classifiers using colour information in order to segment the nuclei and cytoplasm of megakaryocytes in H&E slides with success rates of up to 95% for areas of high confidence. Although the work achieved high pixel accuracy, it must be noted that abnormally shaped MPN megakaryocyte clusters were not tested and manual annotation of the slides is required to establish ground truth colours. Song *et al.* [16] trained classifiers to detect cytoplasm and nuclei of abnormal H&E stained megakaryocytes located close to other cells, achieving a mean accuracy of 80% measured by the Dice coefficient (see Appendix C for an explanation of this). Again, this paper did not extract any criteria for differential diagnosis of MPNs.

2.4 Research Gap

There has been continued debate on the application of existing WHO criteria to differential diagnosis of ET and PMF focused around the key indicators of megakaryocyte clustering and morphology. Studies on inter-observer reliability show that manual identification from bone marrow images can be unreliable without extensive cross-validation, which is not always done by practicing pathologists. No definitive quantitative criteria have yet emerged for this problem.

Previous investigation into using computer vision techniques for this area

was done by Ballarò *et al.* [15] with relatively high classification sensitivity of 90%+. However, there were a number of gaps in this study that are open for improvement. Firstly, it was performed using a dataset of specially selected mature, separated well-imaged megakaryocytes that are atypical of a real-world dataset. They only classify megakaryocytes individually despite real patients having megakaryocytes indicative of multiple disease states, and did not consider megakaryocyte clustering or population features for each patient, considered key features in the WHO criteria [1]. They also do not quantify the confidence intervals for their features. No automated method yet exists for the differential diagnosis of ET and PMF using routine bone marrow samples.

Clearly, a gap exists in the research into diagnosing ET and PMF. This thesis fills that gap by developing a semi-automated method for the segmentation and classification of bone marrow images into normal, ET, pre-fibrotic PMF and PMF categories, investigating clustering in the various disease states and characterising the analysed features.

CHAPTER 3

Design and Implementation

In this chapter the methodology used to fill the research gap identified in the Literature Review (Chapter 2) is explained. The overall methodology can be split into three stages, as seen in Figure 3.1.

First, a convolutional neural network (CNN) is used to segment the raw bone marrow images into labelled regions of interest that may contain a megakaryocyte. Since these regions can contain segmentation errors due to the variations in the input, a secondary machine learning algorithm is used to fine-tune them into clear labelled regions of megakaryocyte nucleus and cytoplasm.

Next, the pixels are grouped into objects and feature vector is computed for each megakaryocyte. The features are based on a mixture of shape and size features, similar to those from Ballarò *et al.* [15]. One of the goals of this project was for the results to be practical and understandable by pathologists, so quantitative numeric features were extracted instead of using a neural network to learn features that could not be replicated by pathologists. These features are also aggregated for each patient into a set of patient-level bone marrow features, similar to what a pathologist would use for a diagnosis. A number of clustering algorithms are used to group megakaryocytes into different types of clusters based on their position. The resultant cluster features are grouped by patient and also analysed separately.

Finally, both patient and individual megakaryocyte feature vectors are put into datasets for classification and criteria analysis. Different classifiers are trained on the data using ground truth diagnoses from the Translation Cancer Pathology Laboratory. Decision tree classifiers are used to obtain rules that pathologists can use to replicate the final results. The full results of this methodology are explained in Chapter 4.

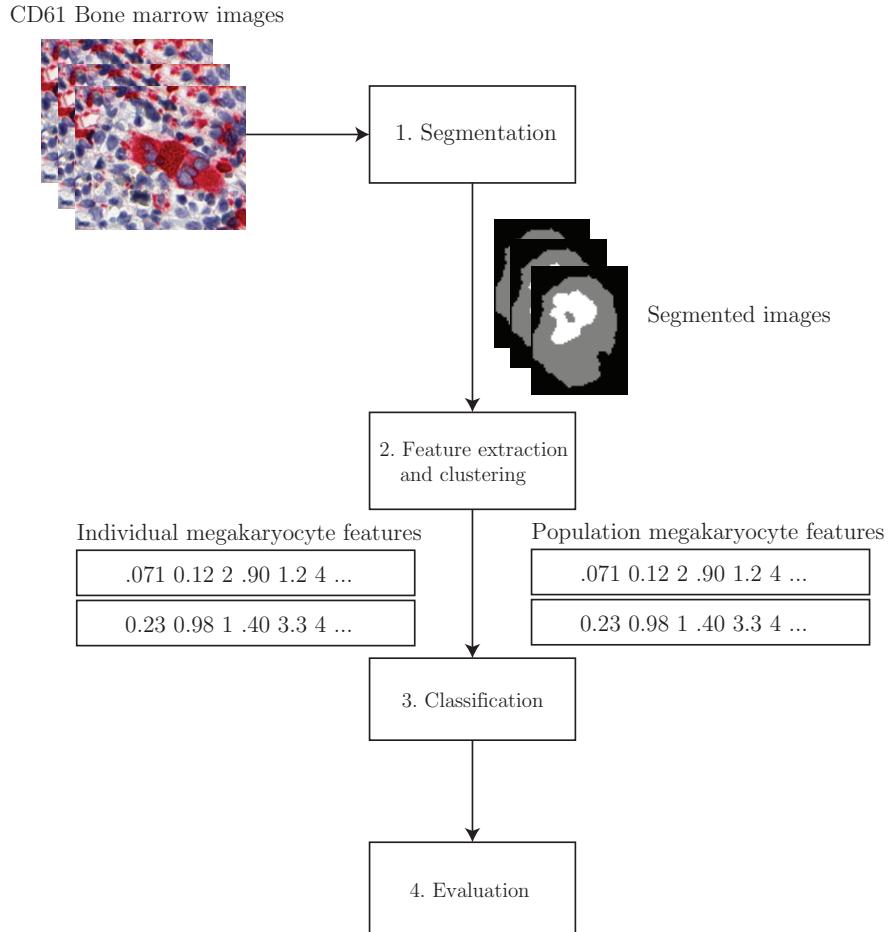


Figure 3.1: A flow diagram of the overall system structure.

3.1 Data Preparation

3.1.1 Input Data Selection

As seen in Chapter 2, two possible imaging techniques were identified, confocal and optical microscopy, with the primary tradeoff being resolution and potential features with unfamiliarity. Although it is clear that confocal microscopy as presented in Acar *et al.* [18] achieves both high resolution and access to three-dimensional features, closer examination revealed a number of problems. Firstly, the use of such a relatively novel technique means that no pre-existing MPN image datasets exist, necessitating collection of a new dataset. Due to not having all of the equipment and antibodies used by Takaku *et al.* [13], developing an alter-

native staining method to collect images proved difficult. Secondly, the confocal microscopy equipment available was unlikely to image a volume more than 100 microns thick, which further reduced the probability of imaging whole clusters of mature megakaryocytes which can range up to 150 microns in thickness. In spite of confocal microscopy's relative advantages, optical imaging was chosen due to the existence of previously collected data and greater familiarity to histologists.

Traditionally, MPNs are sectioned and stained using hematoxylin and eosin (H&E), which highlight the cytoplasm and nuclei of all cells for pathologists to use in differential diagnosis. For this reason, previous segmentation work in this area such as Ballarò *et al.* [15] has focused on analysis of H&E images. Although megakaryocytes can be identified relatively easily under this stain by pathologists, confusion can arise when analysed by a computer because all cells share a common stained colour and can be close together, as seen in Song *et al.* [16].

In order to circumvent this problem and provide the classification algorithms with the best possible segmentation results, a dataset of CD61-stained images was used, collected by the Translational Cancer Pathology Laboratory [19], [20]. CD61 staining gives a high, readily identifiable contrast in colour between megakaryocyte cytoplasm and the background stain, allowing segmentation algorithms to better segment the resulting images, even at low resolutions.

The dataset included a mixture of Tissue Microarray (TMA) images and bone marrow sections. TMAs are essentially a ‘biopsies of biopsy’ approach: first, many small biopsies are taken from a larger bone marrow biopsy and then sectioned and imaged all at once. The resultant images still show clusters of megakaryocytes but are smaller than a full bone marrow section (1.5mm circular TMAs vs 1cm sections of bone marrow), as shown in Figure 3.2. Information about megakaryocyte clustering close to the bone is impossible to know in a TMA, since the TMA cores are subsets of the entire bone marrow biopsy.

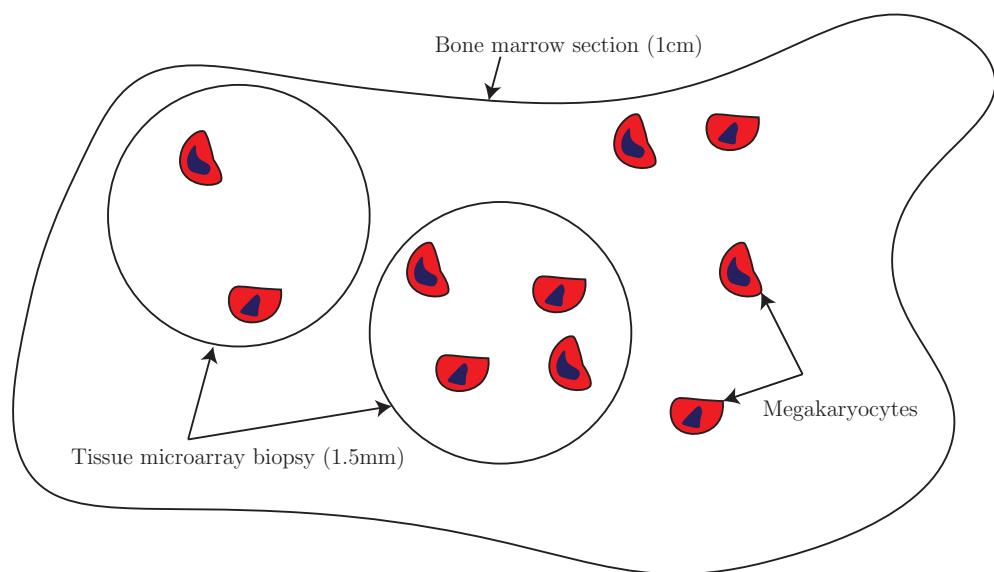


Figure 3.2: A visualisation of the Tissue Microarray approach. Multiple TMAs from the same section are shown for illustrative purposes but in the final dataset only a single TMA from each section was taken.

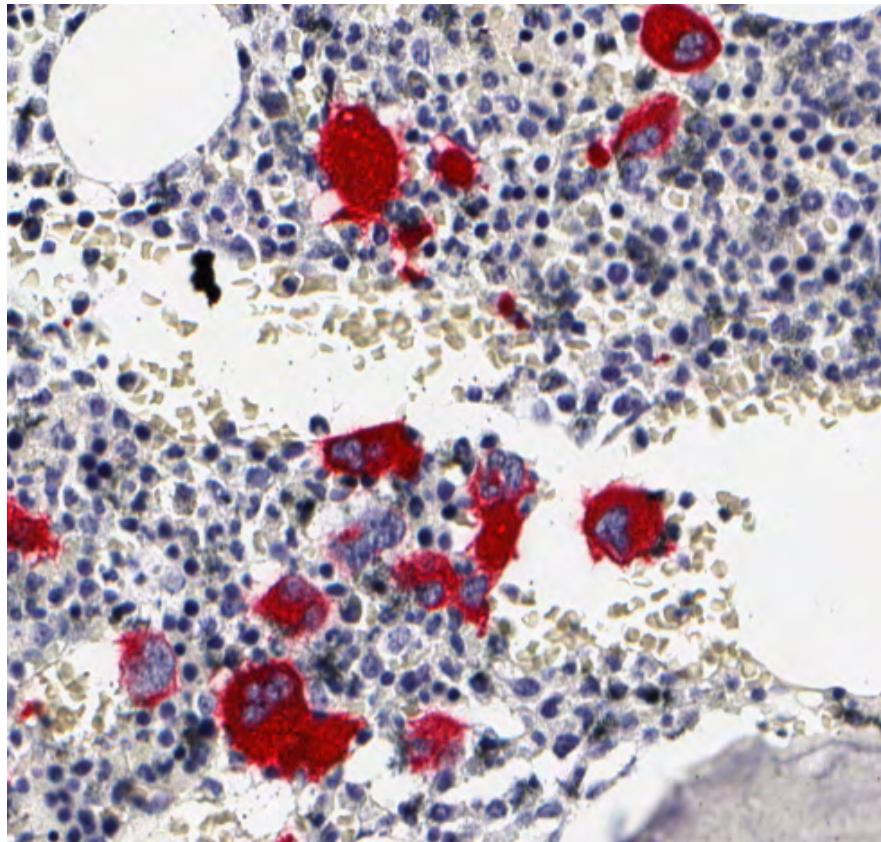


Figure 3.3: Sample CD61-stained bone marrow tissue. Megakaryocyte cytoplasm is clearly visible in red with nuclei in blue. Also note the background tissue consisting of white fat cells, leukocytes, granulocytes and other matter.

The diagnoses for the samples were also provided by the Translational Cancer Pathology Laboratory. Multiple pathologists used the WHO criteria for their diagnosis [20], and categorised the samples into ET, PMF, normal and post-ET PMF. No grading of PMF fibrosis (as described in Section 2.1.1) was performed by the pathologists, so an additional pre-fibrotic PMF classification was added by hand, as described in Chapter 4.

3.2 Segmentation

In contrast to the previous work in the field [15]–[17] a two-stage segmentation algorithm was used: firstly, a convolutional neural network (CNN) for region of interest detection, followed by a machine learning segmentation algorithm for detailed pixel labelling. There were a number of reasons for this choice:

1. *Robust to variations*: the previous algorithms in this field are specific to H&E staining and even more so the dataset used, especially in Ballarò *et al.* [15] which used manually filtered and centered images of megakaryocytes. The algorithm used here is intended to be robust to changes in megakaryocyte clustering and to work on entire sections of unknown bone marrow instead of artificially segmented regions. CNNs are able to recognise higher-level features in the images which allows it to work with different stain intensities.
2. *Flexible to detail*: a two-stage system was used due to the low level of detail in some of the imaged megakaryocytes. The CNN was able to detect interesting regions for the machine learning segmentation algorithm to label. This cut down on training time while still capturing fine detail.
3. *Future adaptability*: both segmentation stages require little dataset and domain-specific knowledge before they are effective at labelling pixels. The CNN can be retrained for any future dataset or feature set simply by using new training data and improved over time with general-purpose developments in the field. In contrast, the algorithms described in Ballarò *et al.* [15] and Song *et al.* [16] used image-specific thresholds and disease state-specific models of megakaryocyte contour shapes to achieve good segmentation results, leaving them less open to future adaptation.

CNNs usually require large training datasets to avoid overfitting to any specific dataset in the image. This was mitigated by using pixel-level information instead of megakaryocyte image information, greatly increasing the number of training samples in the dataset.

Generally the use of CNNs is performance-limited since training, testing and running can be quite time-intensive. However the tested CNN took a maximum of 30 minutes per bone marrow image to run on consumer-grade hardware, well under the typical time needed to stain and prepare bone marrow tissue. For this reason, CNN performance was not taken into consideration for this project.

3.2.1 Convolutional Neural Network

CNNs are conceptually very simple, but have been used to great success in other fields for image recognition. In general, modern CNNs consist of a stack of convolutional layers ending with a fully-connected and loss layer. The convolutional layers have three sub-layers: a filter layer, where each neuron outputs a convolution across a volume of the input, a pooling layer, which downsamples the

output of the filter layer to prevent overfitting, and a rectified linear unit (ReLU) activation layer, which attempts to increase the nonlinearity of the filter layers to avoid the vanishing gradient problem. The fully-connected layer is placed last after the convolutional layers, and is designed to allow each input filter to have varying importance and thus contribution to the classification result. The loss layer is placed last and represents the actual classification result of the network for prediction and backpropagation training.

Although traditionally CNNs are used to predict the class of an entire image at once, for this project per-pixel classification of megakaryocyte cytoplasm and nuclei is required, also known as *semantic segmentation*. The pixel labels are used to determine megakaryocyte edges and area, so the final CNN must be able to categorise each individual pixel into one of three classes: cytoplasm, nucleus and background.

Semantic segmentation is at odds with the pooling layers present in whole-image classification CNNs, since they are designed to lose information and narrow the filters down. To combat this the CNN classifies a single pixel at a time, using features of a 20 pixel neighbourhood. The resultant regions of interest are refined into a better segmentation result using the algorithm in Section 3.2.2.

For optimal performance CNN training must be run on the GPU. Caffe, a heavily optimised neural network framework with Python bindings [21], was used in order to avoid writing GPU-specific code. The models were trained and tested using a single NVIDIA GTX980.

The CNN takes as input a feature vector from a 20x20 neighbourhood around each pixel. First, the image is converted into the Hue-Saturation-Value (HSV) colour space - this is important for classification because the hue of megakaryocyte cytoplasm is red in CD61 images. The RGB colour space has undesirable channel gradient variations for small changes in perceived colour, which are common in stained images. A detailed investigation of this can be found in Zarella *et al.* [17].

Next, each channel undergoes a colour histogram normalisation process to reduce noise artefacts in the input which may contribute to overfitting.

Four feature images are computed from each HSV channel of the neighbourhood:

1. The Gaussian blurred neighbourhood. This blurs the input neighbourhood pixels slightly using a Gaussian kernel in order to reduce high frequency features which may otherwise confuse the CNN and cause it to overfit to the data. This feature image also allows the CNN to specifically identify the stain hue of the cytoplasm used, such as red. While this does make the resultant CNN somewhat stain-specific, it is trivial to retrain the network

for alternative colours.

2. The gradient magnitude map of the neighbourhood, taken using a Gaussian derivative. This feature image tells the CNN how the colours are changing across the neighbourhood, which is helpful for determining cytoplasm and nucleus boundaries. Since the background, cytoplasm and nucleus colours are different, this feature helps classify pixels close to the boundary.
3. The neighbourhood filtered using a difference of Gaussians filter. This computes an approximation of the second derivative of the pixel gradients across the neighbourhood and is most useful for highlighting blobs where colour is constant. This feature image can detect higher-level features such as regions of almost constant colour across megakaryocyte cytoplasm and nuclei as well as blank fat cell regions.
4. The neighbourhood after Canny edge detection. The Canny detector is a widely used multi-stage edge detection algorithm that attempts to find minimal edges in the input [22]. This is a more advanced version of the gradient magnitude filter to try and exactly determine edges in the input neighbourhood and increase accuracy.

The feature images are run through the various convolution layers and classified using a softmax loss layer to produce the final pixel classification result. The overall network structure can be seen in Figure 3.4.

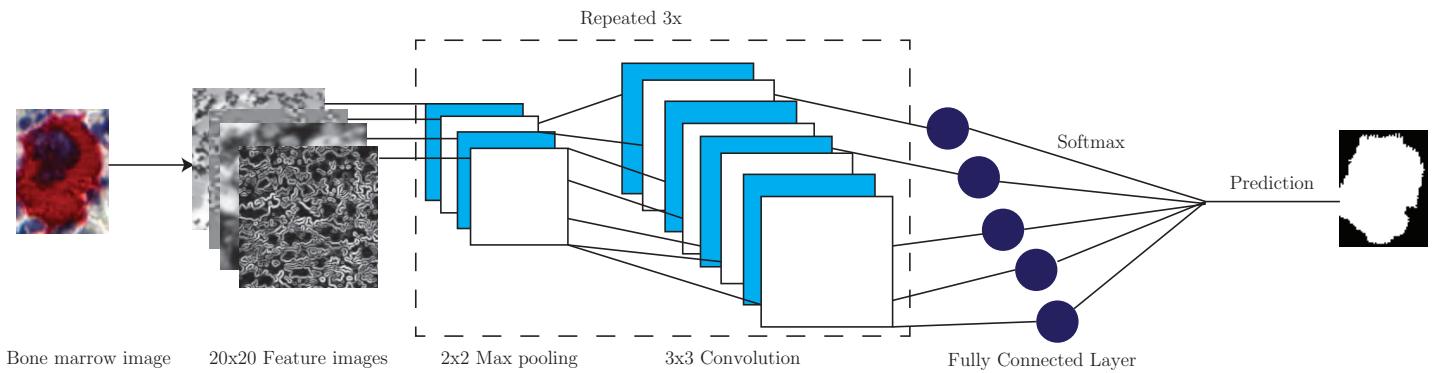


Figure 3.4: A diagram of the convolution neural network

3.2.2 Supervised Classifier Segmentation

The CNN detects nuclei and cytoplasm pixels with high noise and a number of incorrectly labelled pixels due to being trained across a wide variety of input

images. In order to label cytoplasm and nuclei pixels as exactly as possible, a second machine learning segmentation algorithm was used to fine-tune the resultant images.

This segmentation algorithm works similarly to the CNN: a number of feature images are generated per pixel, and a weighted decision tree is used to determine the class of each pixel. The weights of each feature image are learnt using a set of training instances, as explained in Chapter 4. However because the algorithm is running on a smaller image from the CNN, many more feature images can be used without degrading performance, and the decision tree can be quickly retrained with additional examples if a segmentation error is detected.

The classifier software built by Arganda-Carreras *et al.* [23] was used to quickly create accurate training instances from different images. First, a number of 24x24 feature images are generated from the pixel neighbourhood, similar to the CNN. The feature images used are:

- The filtered neighbourhood using the Gaussian blur, difference of Gaussians, Canny edge detection and gradient filters. See Section 3.2.1 for an explanation of these features. 16 different Gaussian blurs are calculated for $\sigma = 1 - 16$, where σ is a parameter describing the standard deviation of the Gaussian kernel (higher σ increases blur).
- Features such as the trace, determinant, orientation and eigenvalues from the Hessian matrix. The Hessian matrix is the matrix of partial second-order pixel derivatives across the x and y domain for the input neighbourhood.
- The sum, mean, standard deviation, median, maximum and minimum of the neighbourhood pixel values.

The feature images are trained using the random decision forest algorithm, first described by Ho [24]. The training instances are split into randomly selected subsets and a forest of decision trees are trained across these subsets instead of the whole dataset. The majority vote from the resultant forest is used to classify each pixel. This reduces classification variance compared to an individual decision tree and increases overall accuracy.

The segmentation algorithm was run manually along the regions of interest to generate high quality labelled megakaryocyte images.

3.3 Feature Extraction

The megakaryocyte segmentation stage outputs a number of image pixels with three different labels associated with a certain pixel value: background, megakaryocyte cytoplasm and megakaryocyte nucleus. The cytoplasm and nuclei of other cells are labelled as part of the background label. These labels are then grouped into megakaryocyte cytoplasm and megakaryocyte nuclei objects based on their connectivity.

3.3.1 Megakaryocyte Labelling

First, sets of connected pixel labels must be labelled into different megakaryocytes. The connected components algorithm was used to group pixels with matching labels, which simply labels pixels according to their adjacent neighbours if they are the same colour. Some additional conditions were applied to avoid incorrectly labelled megakaryocytes:

- Each megakaryocyte cytoplasm label must be 1-connected whereas for the megakaryocyte nucleus this restriction is loosened to 2-connected. This minimises the edges of the grouped cytoplasm since overlapping megakaryocytes are expected to have overlapping cytoplasm more frequently.
- The nucleus of a megakaryocyte must be connected to cytoplasm and each cytoplasm must contain at least one nucleus. Although missing nuclei were considered to be important for diagnosis by Wilkins *et al.* [2], it is difficult to tell whether a reasonably sized cytoplasm area is a megakaryocyte without a nucleus given the possibility of incorrect staining or the megakaryocyte being sectioned at the tip instead of through the center of the nucleus. For this reason it was decided to only consider megakaryocytes with nuclei.
- Only megakaryocytes with cytoplasm area at least 200 microns are considered to avoid analysing megakaryocytes that are too immature or potentially incorrectly stained. A typical mature megakaryocyte will have a diameter of around 50-100 microns.

Next, a number of features were extracted to use for ET and PMF classification.

3.3.2 Extracted Features

Ballarò *et al.* [15] list some basic features useful to extract from the megakaryocytes. Additional features were added to come up with the following list:

1. Cytoplasm area
2. Cytoplasm perimeter
3. Nucleus area
4. Nucleus perimeter
5. Total megakaryocyte area
6. Eccentricity: the eccentricity of the best-fitting ellipse around the nucleus
7. Major/minor axis length: the length of the axes of the best-fitting ellipse around the nucleus
8. Nucleus to cell area ratio
9. Number of nuclei
10. Elliptic Fourier measure: this is derived from an elliptic Fourier series approximation to the contour of the nucleus, the methodology for which is described in Kuhl and Giardina [25]. A full explanation of this method can be found in Section 3.3.3.

This feature set is intended to represent the most important morphological changes captured in the WHO criteria [1], namely changes to cell morphology between ET and PMF and the presence of differences in abnormal megakaryocytes. The size differences are captured by the area and perimeter metrics, while the morphological changes are primarily captured by the eccentricity, major/minor axis length and Elliptic Fourier measures. The Elliptic Fourier measure is a much more detailed interrogation of the nucleus contour than the simple ellipse eccentricity. This measure can be averaged across megakaryocyte populations to come up with characteristic MPN shape visualisations.

The resultant features are grouped by patient and MPN classification and stored along with the megakaryocyte centroid for cluster detection.

3.3.3 Elliptic Fourier Feature

The Elliptic Fourier feature first defined in Kuhl and Giardina [25] was used as a way to capture the shape characteristics of abnormal megakaryocytes in a Fourier spectrum. The WHO criteria list only qualitative shape indicators with descriptors such as ‘staghorn-like’ or ‘cloud-like’ megakaryocytes for ET and PMF respectively [1]. The elliptic Fourier algorithm creates a size and rotation-invariant approximation to a closed contour which can be used as a characteristic signature for a particular nucleus or cytoplasm shape.

The overall goal of this algorithm is to utilise the well-known Fourier series function approximation method to approximate a contour and use it as a signature for diagnostic similarity. To perform the Fourier series approximation, the contour is first expressed using a chain encoding, which breaks the contour piecewise into linear segments of arbitrary length (and so resolution). By parameterising the chain code using a random starting point, the spatial derivative of the chain code is transformed into a time derivative for traversing the code. This derivative allows us to solve the Fourier series to find the Fourier series coefficients. The final set of coefficients acts as a signature for the contour and can be made arbitrarily precise by changing the number of coefficients stored. A full explanation of this can be seen in Kuhl and Giardina [25].

The resultant coefficient vector can be made size, translation and rotation independent [25], which makes it ideal for helping to classify the varying nucleus shapes present in ET and PMF.

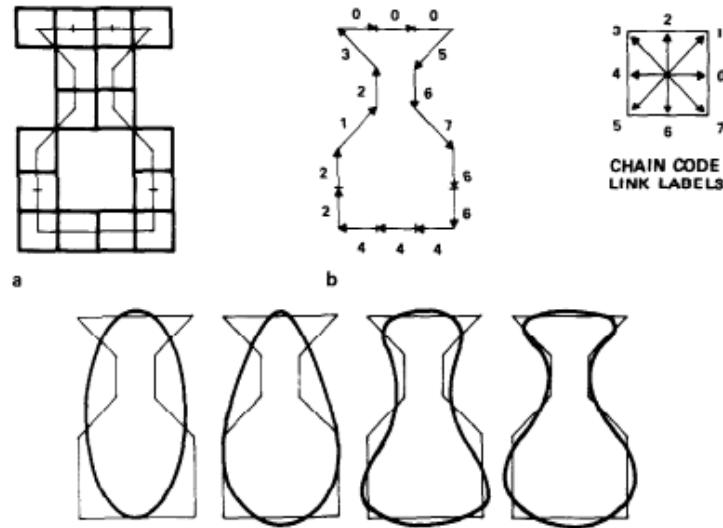


Figure 3.5: A diagram of the Elliptic Fourier Descriptor process, adapted from Kuhl and Giardina [25].

3.3.4 Cluster Detection

A number of cluster-finding algorithms were utilised to group megakaryocytes from a single bone marrow sample into clusters. Each cluster algorithm takes as input megakaryocyte centroids and outputs the labels of the clusters it finds. The megakaryocyte centroids are simply the geometric center of the shape defined by the cytoplasm and nucleus. Since the number of clusters in a patient is unknown, only algorithms that can find an arbitrary number of clusters are considered. Euclidean distance between centroids was used as the primary clustering metric since the other feature vectors were not uniformly distributed. The effect of each clustering algorithm is measured by analysing its impact on patient MPN classification accuracy.

Scikit-Learn [26] was used to implement the clustering algorithms. The tested clustering algorithms were agglomerative clustering, DBSCAN and affinity propagation.

Agglomerative clustering is a simple bottom-up hierarchical clustering algorithm similar to that described in Johnson [27]. Each megakaryocyte begins as a separate cluster. Every iteration, the algorithm merges the pair of clusters that minimise the Ward linkage or the distance variance of the resultant cluster. This is repeated until a maximum distance between clusters is reached.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density clustering algorithm proposed by Ester *et al.* [28] that searches for dense areas. It takes two parameters: a minimum number of points p and a neighbourhood distance d . A megakaryocyte is considered to be a *core point* if at least p are within its neighbourhood distance. Any other megakaryocytes within range of a core point are also connected to the cluster. These parameters must be modified based on the dataset to produce the best result, as explained in Chapter 4.

Affinity propagation is a newer iterative clustering algorithm described by Frey and Dueck [29]. Each megakaryocyte is placed into a fully connected network where each node tries to find the best ‘exemplar’ or cluster representative. Every node communicates to its neighbours with information about the best exemplar it has found and the information it has collected from its neighbours. The algorithm chooses a new set of exemplars from the evidence that each node has collected, and terminates once no changes have been made. It performs better than k-means clustering in terms of minimising cluster variance and does not require specifying the number of clusters [29].

Each clustering algorithm produces a set of cluster labels. The number of clusters and the distance variance of each cluster are used as feature vectors for the classification stage.

3.3.5 Classification and Criteria Analysis

Once the features and clusters have been analysed for the population, the best criteria for classifying a new, unknown tissue sample must be investigated. This process attempts to replicate the final differential diagnosis step that would performed by a pathologist using an automated classifier algorithm,. The analysed features are also characterised across the different classes. The classification methodology is based on comparing multiple standard classifier algorithms trained on both the individual megakaryocyte features and aggregated megakaryocyte features to find the best method of classifying megakaryocytes in the dataset. The features from individual megakaryocytes were expected to result in relatively poor classification rates because of the inherent biological variation in the tissue samples. Each classifier is trained using 10-fold cross-validation.

3.3.6 Classification Algorithms

The classification algorithms were built using Weka [30], which allowed quick comparisons of the various algorithms. Many of the algorithms have additional

parameters that tune the algorithm results: these were left at the default values (i.e. no hyperparameter optimisation occurred).

The first classifier used was the naive Bayes classifier, implemented in Weka using John and Langley [31]. The naive Bayes algorithm makes the assumption that each feature is conditionally independent for a certain class. This assumption means that Bayes' theorem can be used to build up per-class probability estimates for the training feature vectors assuming that they are each independent and normally distributed. Observed feature vectors are classified according to the maximum conditional probability for each class. Clearly, the extracted features such as cell area and perimeter are not independent, so naive Bayes is expected to perform poorly. However this classifier provides a fast baseline prediction that is widely used in classification problems.

The next classifier is the Support Vector Machine (SVM), implemented in Weka as the Sequential Minimisation Algorithm (SMO) proposed in Keerthi *et al.* [32]. This represents each of the features as points in separate planes and trains linear binary classifiers for the entire hyperplane by maximising the margin between the plane and the points in each plane. Since multiple classes exist in this project, a separate SVM is trained for each pair of classes - new feature vectors are tested under each of the pairwise SVMs and the result is the most frequently classified class. This classifier should produce a better result than the naive Bayes classifier since features are not assumed to be independent, but since the features may not be linearly separable it is not expected to be the best performing algorithm.

One of the goals of this thesis is to generate differential diagnosis criteria that are practical, specifically that they can be applied directly by pathologists. The C4.5 (known as J48 in Weka) decision tree classifier [33] was chosen for this reason since it can generate human-readable rules for classification. It generates binary decision trees by iteratively splitting the tree on the feature that maximises information gain, essentially minimising the information entropy of the resultant tree. C4.5 is a greedy algorithm that is very simple and does not result in the best performance but is useful here because it allows pathologists to replicate the final differential diagnosis criteria. For space reasons the final decision tree with over 100 nodes could not be shown in this thesis but it could conceivably be applied directly by pathologists after further research into the most important features.

The K-Nearest-Neighbours (KNN) classifier (known as IBk in Weka) was also used, which simply assigns a class at classification time based on the class of its closest neighbours - each neighbour contributes a vote for its class, and the class with maximum votes is the per-feature classification. This is repeated for each

feature and the most frequent class is the final classification result.

Finally, the multilayer perceptron (MLP) algorithm was tested, which is a neural network with an input layer, some hidden layers and a final output layer. Each neuron activates using a sigmoid function (in this case the hyperbolic tangent from -1 to 1) and a trained weight value. MLPs can model non-linear regression functions and in theory learn higher level concepts than a simple linear classifier.

CHAPTER 4

Evaluation

4.1 Dataset

As outlined in Section 3.1.1, the dataset chosen for this project was bone marrow images stained using CD61 to highlight the megakaryocytes in red. The data collection process and caveats are explained in more detail here.

4.1.1 Image Collection

The dataset was previously collected from patients between 2000-2014 by Path-West Laboratory Medicine (Western Australia) and Queen Mary Hospital (Hong Kong) [19]. The images in the dataset were stained with CD61, an antibody that binds to megakaryocyte cytoplasm, and counterstained with Mayer’s hematoxylin, which stains all cell nuclei in the marrow [19]. This results in an image with blue nuclei for every cell and megakaryocyte cytoplasm in shades of red. The images were taken at 20x optical magnification (0.5 microns per pixel).

The bone marrow was imaged in the form of 96 archived trephine biopsies (sections of bone marrow preserved using either formalin or formalin-mercury fixative [19]), each from a different patient. In order to obtain sections suitable for imaging, a multi-stage process is followed. First the archived bone marrow undergoes acid decalcification to remove any opaque material which would otherwise block the optical microscope and cause issues with sectioning while leaving the marrow relatively intact bar shrinkage. Next, the sample is dehydrated and embedded into a paraffin block. The paraffin block is cut using a microtome and the resultant sections are stained and fixed onto slides for imaging.

Malherbe *et al.* [19] sectioned 1.5mm TMAs at 4 μm intervals, stained them with a number of different antibodies and imaged using a Pixera Pro 600ES microscope. The patients were from a variety of backgrounds, ages, and disease progression, but patients with MPNs had not been medically treated prior to

taking the biopsy [19].

Images from patients with Polycythemia Vera were ignored since they were out of scope for this project. Images that were significantly distorted or had only one or two megakaryocytes were also removed to ensure that good representative undistorted samples for each class were taken.

4.1.2 Diagnosis Collection

The diagnoses for the dataset were also provided by PathWest Laboratory Medicine. Multiple pathologists followed the WHO criteria for diagnosing each individual patient [19] and put each patient into five classes: normal (control), ET, post-ET PMF, PV and PMF (Table 4.1). Inter-observer reliability was not recorded in the dataset, so some disagreement is unavoidable. No description of disease progression was given, so post-ET PMF and PMF were recategorised into pre-fibrotic and post-fibrotic PMF. The WHO criteria include a quantitative grading system for PMF based on median values for various indicators such as spleen size, haemoglobin levels, and thrombocyte frequency as well as a semi-quantitative four-grade system based on bone marrow fibrosis evident in the images. As this thesis did not have access to patient medical records the recategorisation was based on whether fibrosis was visible or not visible in the sample images (in the WHO criteria, this is the difference between PMF grade 0-1 and grade 2-3).

Bone marrow fibrosis is the formation of excess fibrotic binding tissue in the marrow. In PMF it manifests in an increase in connected reticulin fibres and collagen bundles, with more prevalence throughout the bone marrow as the disease progresses. Many of the difficulties in differentiating between ET and pre-fibrotic PMF occur because fibrotic tissue is not as prevalent. An example of visible fibrosis and normal tissue can be seen in Figure 4.1.

The PV images were discarded from the dataset, since relatively strong diagnostic criteria exist for this disease state.

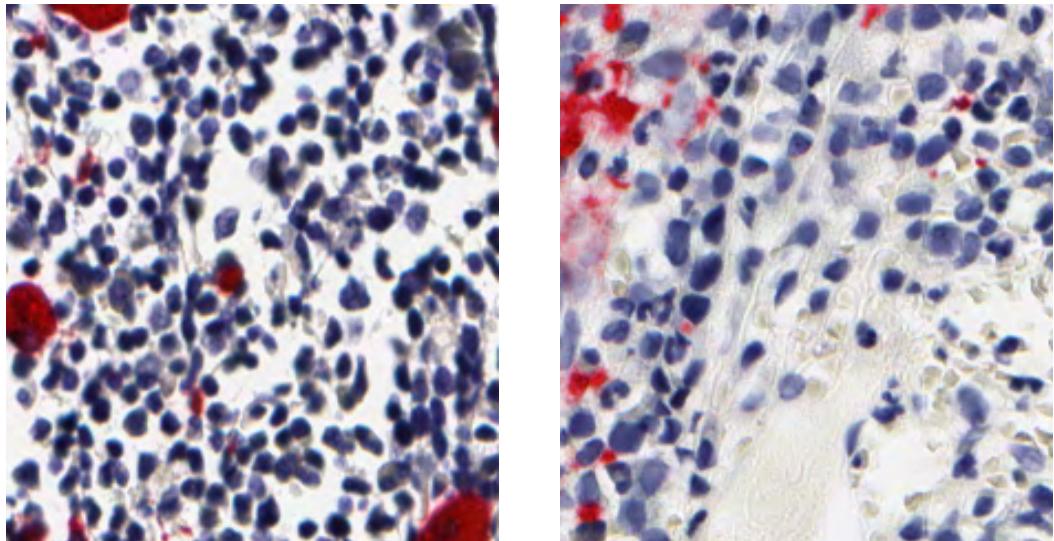


Figure 4.1: Pre-fibrotic PMF tissue on the left compared with fibrotic PMF tissue on the right. Fibrosis has manifested in the grey collagen tissue that obscures parts of the image on the right. There is no evidence of widespread reticulin fibres or collagen in the tissue on the left.

Class	Number of images
Normal	12 TMAs
Essential thrombocythosis	40 TMAs
Pre-fibrotic primary myelofibrosis	17 TMAs
Primary myelofibrosis	14 TMAs

Table 4.1: The final image dataset, grouped by diagnosis

4.1.3 Segmentation Training Set

A ground truth is required for categorising each pixel into nucleus, cytoplasm and background. Megakaryocytes were identified in 10 images across the different classes and painted their cytoplasm and nuclei in black and white respectively for use in the training set. Training pixels for each class were taken from different tissue samples and tissue types to create a representative training dataset.

The final pixels used for training were taken randomly from the overall dataset to both limit training time and avoid overfitting effects over multiple training runs. In the training phase, 30000 pixels were used with 80% in the training set

and 20% in the testing set. A separate validation phase with 10000 pixels was used to test the accuracy of the trained CNN.

4.1.4 Caveats

4.1.4.1 Image Defects

The bone marrow staining, sectioning and imaging process is not perfect. Many artefacts may be introduced into the resultant images, ranging from severe defects to minor distortions and inconsistently stained areas. Some examples of various errors in the dataset can be seen in Figure 4.2.

In general, the dataset errors can be categorised into recoverable and unrecoverable artefacts. An unrecoverable artefact in a particular image means that it must be cropped to remove the undesirable areas. Some examples are Figure 4.2(a) and (b). If the defect covers enough of the image area, it must be discarded entirely. Generally, this type of artefact is easier to deal with since the offending areas can simply be removed.

A recoverable artefact is something that must be worked around in the feature extraction stage because the underlying images still contain good megakaryocyte data, just distorted. Examples include stain intensity variations, different counter-stain colour or platelet ribbons stained as if they were megakaryocyte cytoplasm. Most of these are fixed by restricting detected objects to 50-200 microns and colour normalisation to combat the stain colour variations. However because platelet ribbons can be almost megakaryocyte-sized and obscure the main cytoplasm, the algorithm was forced to require the existence of a nucleus within each megakaryocyte to avoid the possibility of mistakenly extracting platelet features as opposed to megakaryocyte features.

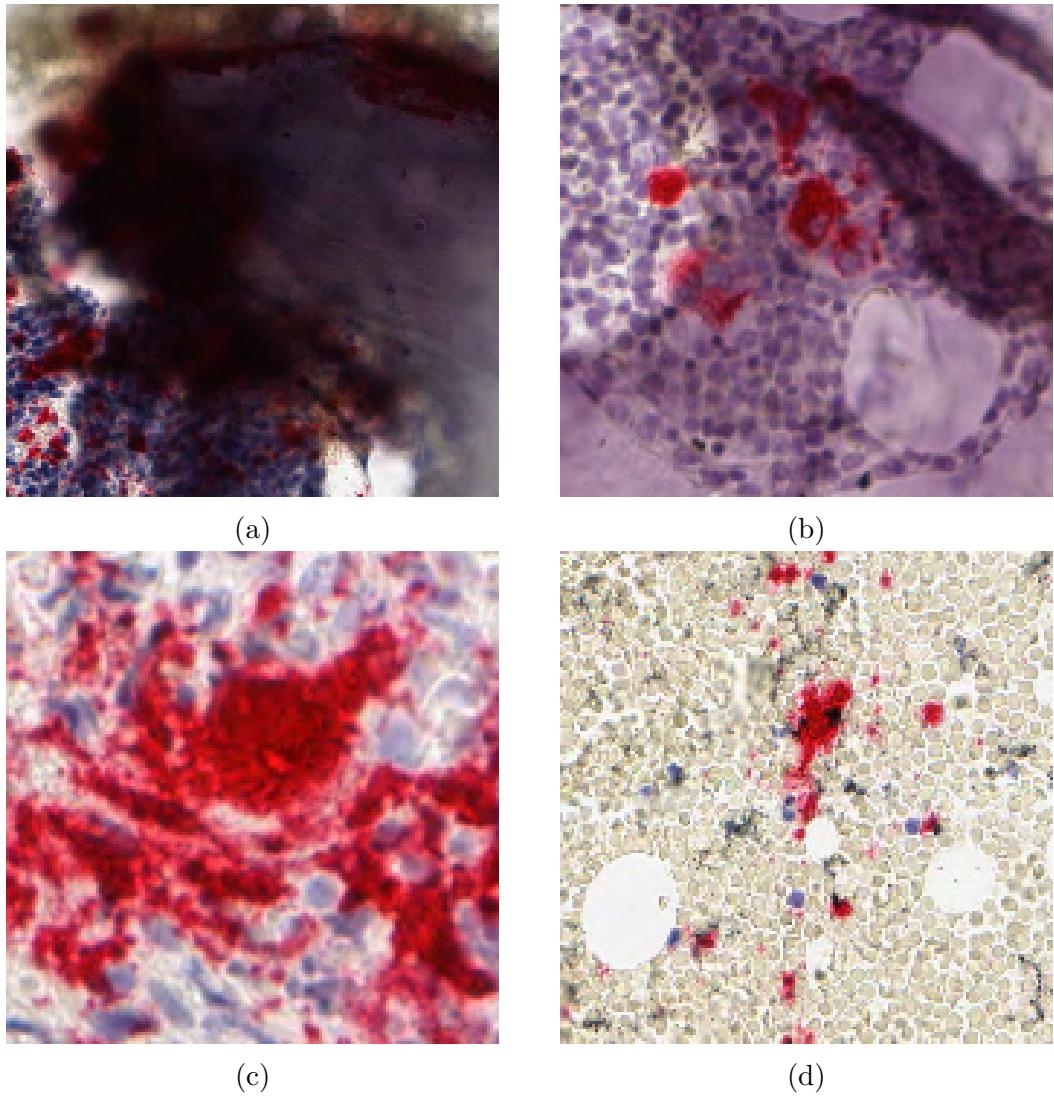


Figure 4.2: Sample image defects from the dataset

(a) Severe discolouration and damage to the tissue sample. (b) Debris obscures and changes the megakaryocyte nucleus colour. (c) Partial staining of platelets close to a megakaryocyte. (d) Distorted background tissue from a PMF patient, requiring colour normalisation.

4.1.4.2 Training Set

The training pixels were manually segmented to provide a ground truth for the two segmentation algorithms. However, it is difficult to see from raw pixels the exact edge of a megakaryocyte, especially if it borders another megakaryocyte.

For this reason, a bias exists in the training data set towards pixels clearly inside the megakaryocyte nucleus and cytoplasm. Thus, some fuzziness in the edge detection of megakaryocytes is expected, but since it is applied uniformly it should not adversely impact the feature extraction phase.

4.1.4.3 Diagnoses

The diagnoses from Malherbe *et al.* [19] did not contain inter-observer reliability indicators or PMF grading diagnosis, which negatively impact the accuracy of the ground truth dataset. Unfortunately, no more information could be collected from the pathologists so the PMF grading was done manually. The main uncertainty in the diagnoses thus lie in the distinction between pre-fibrotic PMF and ET as pathologists find it hard to differentiate these two diseases reliably as explained in Chapter 2. On the other hand, normal bone marrow and advanced PMF marrow is easier to identify for pathologists so good diagnoses are expected for these disease states. Future studies should rely on more reliable diagnoses, but for this project only the data from the Translational Cancer Pathology Laboratory was available.

4.2 Segmentation

4.2.1 Training

The CNN seen in Figure 3.4 was trained on a training dataset with 30000 pixel examples per class using the Caffe neural network framework [21]. A Stochastic Gradient Descent (SGD) solver was used to semi-randomly explore the search space. The SGD solver optimises the network by updating the weights for back-propagation such that the gradient of the loss function of a random input sample is minimised and thus accuracy is maximised over time. The network was tested every 1000 iterations against the full training set to compare accuracy values across different iterations. The most accurate Caffe model from the training phase was to produce input regions of interest for the machine learning classifier.

The machine learning classifier was again trained using manually labelled data from each region of interest.

4.2.2 Results

A summary of the segmented images in terms of number of megakaryocytes can be seen in Table 4.2. First, the CNN separated the raw bone marrow images into a number of regions of interest. Many of these were incorrectly or partially segmented. Next, the machine learning classifier was used for high quality segmentation of the regions of interest. Since the second step involved manual verification to avoid segmentation errors, not all of the bone marrow images could be fully segmented. As the clustering algorithms rely on all or most of the megakaryocytes in a TMA image being segmented, not all of the TMAs could be used for cluster analysis.

Class	CNN		Machine learning classifier	
	TMAs	Regions of interest	TMAs	Megakaryocytes
Normal	12	1074	6	97
ET	40	5449	3	64
Pre-fibrotic PMF	17	1318	3	73
PMF	14	4156	2	137

Table 4.2: Segmentation results - the number of TMAs represents the number of TMAs that have been completely verified and segmented

Pixel-level segmentation accuracy for the machine learning classifier could not be determined across the dataset since the final segmentation was megakaryocyte specific and often included megakaryocytes that had to be excluded from analysis due to overlapping, having missing nuclei or the other caveats mentioned in Section 4.1.4. However, each image was manually verified to a high standard.

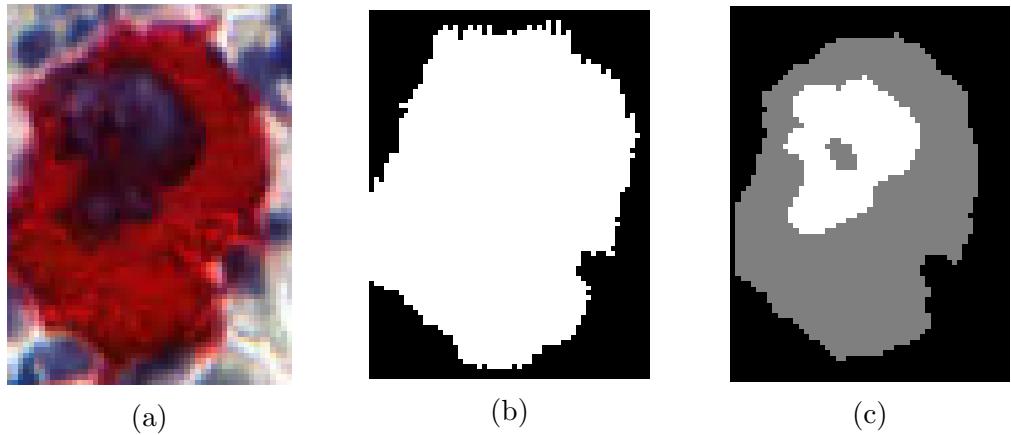


Figure 4.3: A sample segmentation of a well-formed megakaryocyte showing the two segmentation stages. Black, grey and white pixels represent background, cytoplasm and nucleus pixels respectively.

(a) The raw megakaryocyte to segment, in this case from pre-fibrotic PMF. (b) The region of interest found by the convolutional neural network. (c) The final supervised learning classifier segmentation

4.3 Feature Extraction

4.3.1 Extraction

The labelled pixels from the segmentation stage are grouped and transformed into a feature vector, as outlined in Section 3.3.2. The number of pixels in both the nuclei and cytoplasm are counted to determine area and perimeter. The ratio between the nuclei area compared to the total cell area is recorded. An ellipse is fitted to both the largest nucleus and the cytoplasm to measure its eccentricity and major and minor axis length. Finally, the Elliptic Fourier algorithm [25] is used to determine the Fourier coefficients of the nucleus and cytoplasm shape. Each of the coefficients is used as a separate ‘feature’, since they are rotation, translation and size invariant.

4.4 Classification and Criteria Analysis

4.4.1 Machine Learning algorithms

In order to learn the relationship between features extracted feature vectors were fed along with their class into a number of supervised learning classifiers as explained in Section 3.3.6.

Two different types of feature vectors were tested: the individual megakaryocytes feature vectors and aggregated feature vectors per TMA. The aggregated feature vectors are simply the mean of all megakaryocyte features for a particular TMA. There were relatively few TMA samples which give complete megakaryocyte population and clustering information. For this reason, resampling in the form of bootstrapping was employed in order to broaden the sample space give enough statistical descriptors to the classifiers to properly classify megakaryocyte populations. Clustering information cannot be resampled since megakaryocyte clustering is dependent on the patient’s biology and disease progression: instead, the clusters that were extracted are analysed separately.

Bootstrapping is a powerful technique which refers to generating synthetic datasets by resampling with replacement from a set of samples [34]. The resultant datasets help define the original distribution from which the real sample was taken if enough synthetic samples are created [34]. In this case there are plenty of sample megakaryocytes from each class but few composite megakaryocyte sample populations. 5000 resampled megakaryocyte populations of 15 megakaryocytes for each of the classes are generated to test classification performance. 15 megakaryocytes per resampled population are used because there are 15 megakaryocytes on average per normal TMA, so it is representative of a new patient.

This resampling method results in some artificial increases in classification accuracy since the features may be resampled multiple times and introduce bias towards the most prevalent values while reducing variance since the mean of each resampled population is used. In order to quantify and attempt to mitigate this, the bootstrap and classification process is repeated 200 times in order to generate confidence intervals for the accuracy measurements. 95% confidence intervals are measured using the bootstrap method with 10000 resamples.

4.4.2 Classifier Results

A number of performance statistics were recorded for each classifier. An explanation of these statistics can be found in Appendix C. Briefly:

- **Accuracy:** the percentage of the input that was correctly classified.
- **Root mean squared error (RMSE):** the average standard deviation of the difference between the observed values and the predicted values for each feature. Note that this is unitless in this case since the prediction result is simply 1 or 0 for a correct and incorrect result. In this case a low RMSE is closely linked to accuracy, since the classifier is working with nominal instead of numeric classes that would normally be used to refine the RMSE.
- **κ -score:** a measure of accuracy after correction for random predictions. Values below zero represent correct predictions that could be purely down to chance. As the score increases to 1, agreement is considered to be perfect and unrelated to random predictions.
- **F-score:** another measure of accuracy that calculates the proportion of true positives to false positives and false negatives. The F-score ranges from 0 to 1, with 1 being perfect accuracy.

Classifier	Accuracy (%)	RMSE	κ -score
Naïve Bayes	42.05	0.51	0.20
KNN	34.50	0.56	0.09
C4.5	34.50	0.55	0.09
SVM	34.5	0.43	0.13
Multilayer Perceptron	38.01	0.50	0.14

Table 4.3: Classification accuracy from each algorithm on individual megakaryocyte features

Classifier	Accuracy (%)	RMSE	κ -score	F-score
Naïve Bayes	76.81-76.98	0.30-0.30	0.69-0.69	0.77-0.78
KNN	88.06-88.21	0.21-0.21	0.84-0.84	0.88-0.88
C4.5	85.38-85.57	0.26-0.26	0.81-0.81	0.85-0.86
SVM	91.41-91.53	0.32-0.33	0.89-0.89	0.91-0.92
Multilayer Perceptron	93.41-93.61	0.16-0.16	0.91-0.91	0.94-0.94

Table 4.4: Classification accuracy from each algorithm on the bootstrapped megakaryocyte populations. The values shown are of 95% confidence interval values of the mean.

Feature	Normal	ET	PMF	Early PMF
Cytoplasm perimeter (microns)	85.35-98.69	77.06-88.70	85.85-97.04	87.12-96.03
Cytoplasm area (microns ²)	759.23-994.09	542.99-697.92	619.88-785.05	637.99-768.22
Nucleus area (microns ²)	120.20-177.22	99.10-145.01	157.05-196.48	110.29-139.27
Nucleus perimeter (microns)	27.50-36.29	25.91-33.54	33.11-39.12	26.06-30.78
Nucleus to cell area ratio	0.13-0.17	0.15-0.20	0.21-0.26	0.14-0.17
Nucleus eccentricity	0.64-0.71	0.70-0.77	0.75-0.80	0.68-0.73

Table 4.5: 95% confidence intervals for the mean of each megakaryocyte feature (Elliptic Fourier coefficients omitted)

4.4.3 Clustering Results

As seen in Table 4.2, only a few TMAs per class had every single region of interest segmented. Clustering information can only be accurately determined for TMAs that have been completely segmented in this fashion since it is otherwise impossible to exactly determine the cluster centroid if some megakaryocytes have been omitted. Due to the low sample size of clusters, clustering information could not be used as an input for the machine learning classifiers. Resampling could not be used to generate new virtual clusters since the resultant clusters are unlikely to be biologically plausible.

Instead, the clustering algorithms were used to characterise the clusters in the data to create a methodology that could be used in future work. As explained in Section 3.3.4, three different clustering algorithms were tested on the centroids of the segmented megakaryocytes. The Silhouette Coefficient was used to compare the effectiveness of each clustering algorithm in separating the centroids into different clusters.

The Silhouette Coefficient was first proposed by Rousseeuw [35] and measures the similarity of each cluster against the separation of different clusters. A score of -1 represents poor, incorrectly assigned clusters whereas a score of 1 represents perfectly assigned clusters. A score close to 0 represents overlapping clusters. The Silhouette Coefficient is calculated using the following formula for each cluster, with a being the average intra-cluster distance and b being the average distance to the nearest separate cluster:

$$\frac{b - a}{\max(a, b)}$$

The average Silhouette Coefficient across all of the clusters and classes for each clustering algorithm is shown in Table 4.6. Since DBSCAN takes the max-

imum neighbourhood size as a parameter, the neighbourhood size that gave the maximum Silhouette Coefficient for each class was used to obtain a representative Silhouette Coefficient.

Clustering algorithm	Average Silhouette Coefficient
Agglomerative clustering	0.496
Affinity propagation	0.501
DBSCAN	0.450

Table 4.6: Average Silhouette Coefficients for each clustering algorithm

The best Silhouette Coefficient was found using the affinity propagation algorithm. The clusters found using this algorithm are characterised in Table 4.7.

Class	Intra-cluster distance (microns)	Clusters per TMA
Normal	120.78	~3.7
ET	70.06	5.0
PMF	66.81	6.0
Pre-fibrotic PMF	89.21	6.0

Table 4.7: Mean cluster statistics for each class using the affinity propagation clustering algorithm

4.5 Discussion

4.5.1 Segmentation

Overall, the segmentation algorithm was successful in that the CNN detected regions of interest that were refined using a machine learning segmentation algorithm. However, the algorithm could not reliably be used to automatically segment an entire bone marrow image.

Initially this seems strange since previous automated segmentation work done by e.g. Ballarò *et al.* [15] and Song *et al.* [16] achieved almost perfect segmentation accuracy. However, a number of differences in the dataset for this thesis compared to theirs precluded fully automatic segmentation.

The primary issue faced was that in order to obtain accurate clustering information, most if not all of the megakaryocytes per TMA needed to be identified

and segmented. This presents a more difficult problem than that tackled in previous works which segmented only mature, separated megakaryocytes. Advanced disease states such as PMF often manifest elongated, overlapping nuclei and cytoplasm, which is a clear signal for a pathologist but is difficult for an automated segmentation algorithm to reliably separate into distinct clustered megakaryocytes.

Secondly, there were wide nuclei colour variations in the image data. The cytoplasm of each megakaryocyte was readily identifiable by the algorithm due to its predominantly red nature but a mixture of colour variations and low resolution meant that the nuclei pixels were hard to reliably identify. For an example of both of these problems, see Figure 4.4.

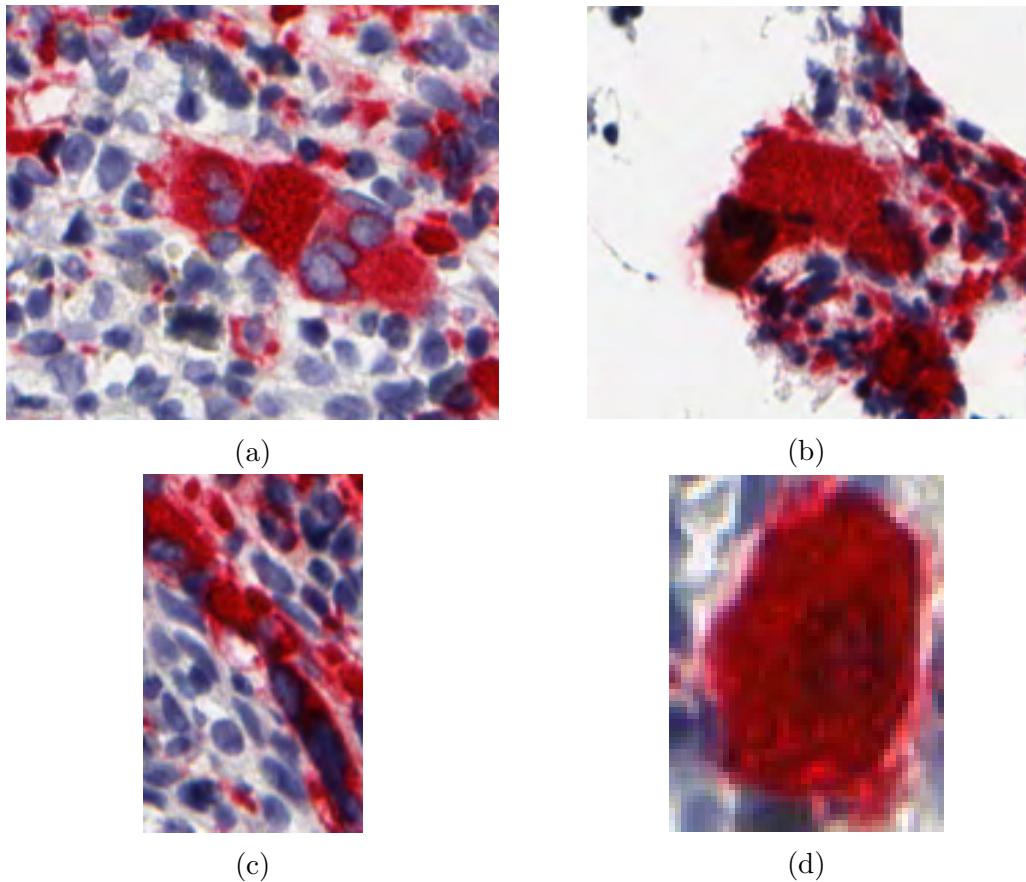


Figure 4.4: Sample megakaryocytes that were incorrectly classified and necessitated manual segmentation.

(a) Three overlapping megakaryocytes (b) Two overlapping megakaryocytes and overlapping granulocytes (c) Elongated cytoplasm and nuclei in PMF (d) A poorly stained nucleus

These issues meant that although well-formed mature megakaryocytes were easy to segment, it was impossible to automatically segment all of the megakaryocytes in the samples without verification. Instead each megakaryocyte had to be segmented and verified individually using the machine learning classifier, a time-consuming process that limited the number of TMAs with clustering information.

Each megakaryocyte had to be examined for around 1-5 minutes to correct segmentation errors, mirroring what a pathologist typically does in the differential diagnosis stage to examine megakaryocyte morphology. Although this manual segmentation process is a limitation of this project, it is also something that can be actively improved with a mixture of better tissue preparation, megakaryocyte-specific staining techniques, higher magnification images and a larger ground truth database for the segmentation algorithms. Once the segmentation process is complete the classification algorithm is able to automatically extract features that a pathologist cannot such as the Elliptic Fourier descriptors and compare them to previously seen data.

4.5.2 Classification

For individual megakaryocytes, the classification results were quite poor with a maximum of 42.05% accuracy, well below the roughly 90% per-class classification accuracy in Ballarò *et al.* [15]. Furthermore, this maximum accuracy came from the Naïve Bayes classifier, which assumes that each feature is independent, suggesting that the classifiers have failed to capture the relationship between each feature. This assessment is confirmed by the low κ -scores across each classifier, which means that the classifiers often perform little better than random chance. However, this result was actually expected given the differences in dataset - the segmentation algorithms were used to segment every megakaryocyte in a TMA, whereas Ballarò *et al.* [15] pre-selected mature megakaryocytes that specifically indicated a certain class. Previous work such as Wilkins *et al.* [2] has identified megakaryocytes representative of different disease states in the same patient, such as an ET-like megakaryocyte closely neighbouring a prefibrotic PMF-like megakaryocyte. Since this project tries to segment as many megakaryocyte as possible, even if they are less mature and have less nucleus information, the low classification accuracy is likely an artefact of the dataset choice.

For the bootstrapped megakaryocyte population a much more positive result was obtained, especially with techniques such as the multilayer perceptron obtaining up to 93.41% accuracy. There was some variation in classification results: for example, the naïve Bayes classifier was nearly 10% lower in accuracy than

other results. This shows that the individual features were inter-dependent in some way, since the naïve Bayes classifier assumes independent features.

The multilayer perceptron and the SVM algorithms had the best results, at 91.4% and 93.4% accuracy respectively. Both of these algorithms attempt to find some function that can separate the classes. The high performance of the SVM algorithm which attempts to find linear binary classifiers for each class implies that the megakaryocyte features are relatively separable. Although it could be due to overfitting to the dataset, the multilayer perceptron had the best accuracy which implies that the separation is best represented as a non-linear function.

A comparison of two features can be seen in Figure 4.6. Cytoplasm minor axis length and nucleus area clearly show wide variation in the case of individual megakaryocytes, with some outliers making it difficult to find a clear separation between the different classes. Aggregating these features across megakaryocyte populations reveals that a combination of these two features alone can clearly separate most of the bootstrapped population features. Interestingly, most of the classification errors occur within the pre-fibrotic PMF class, suggesting possible misdiagnosis for the TMAs in this class.

The most important features for classification were total nucleus area and cytoplasm major/minor axis length, found by examining the highest decision nodes in the C4.5 decision tree. The Elliptic Fourier descriptor feature was a worse indicator than the simple major and minor axis length feature, which also approximates shape changes. This was the result of using averaged Elliptic Fourier descriptors instead of considering each megakaryocyte individually: an illustration of this can be seen in Figure 4.5. Although the individual Elliptic Fourier features accurately captured shape features in a translation, rotation and scale invariant manner, a better way to combine these features into a representative shape or set of shapes per population is necessary.

The results were quite consistent over the 200 classification tests, with little difference in the 95% confidence intervals for each classifier. This shows that the classifiers are both consistent and relatively accurate over the input dataset and validates the extracted features and population-based analysis approach. The classifiers could easily be extended with more features such as clustering and more input data to achieve better results.

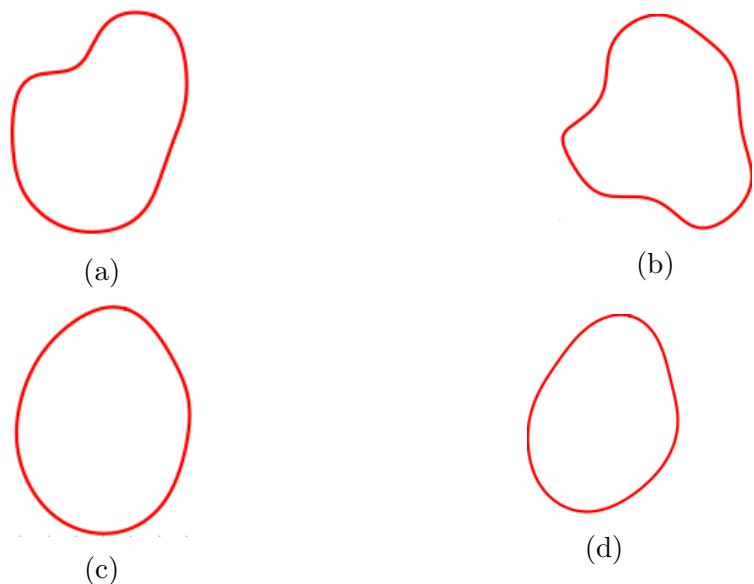
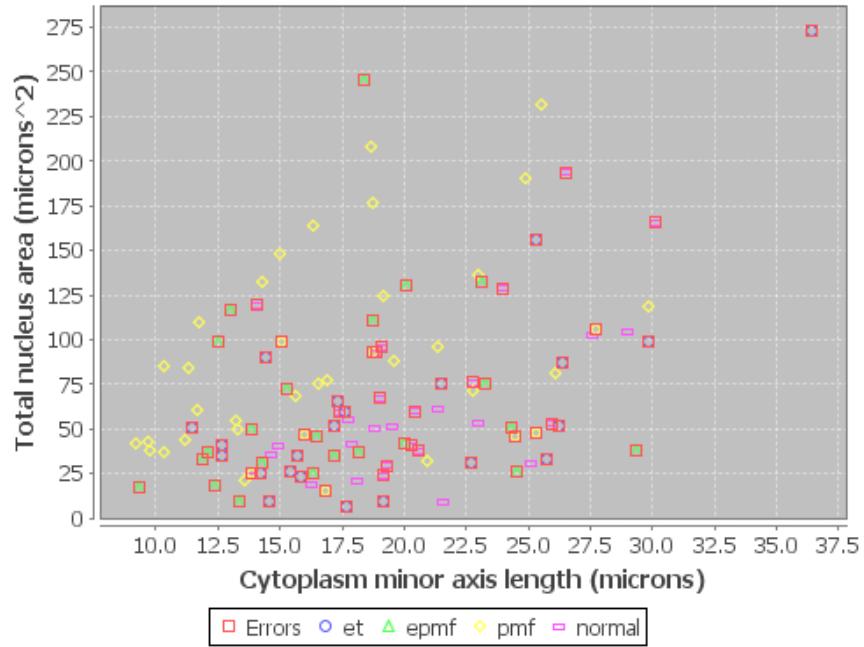
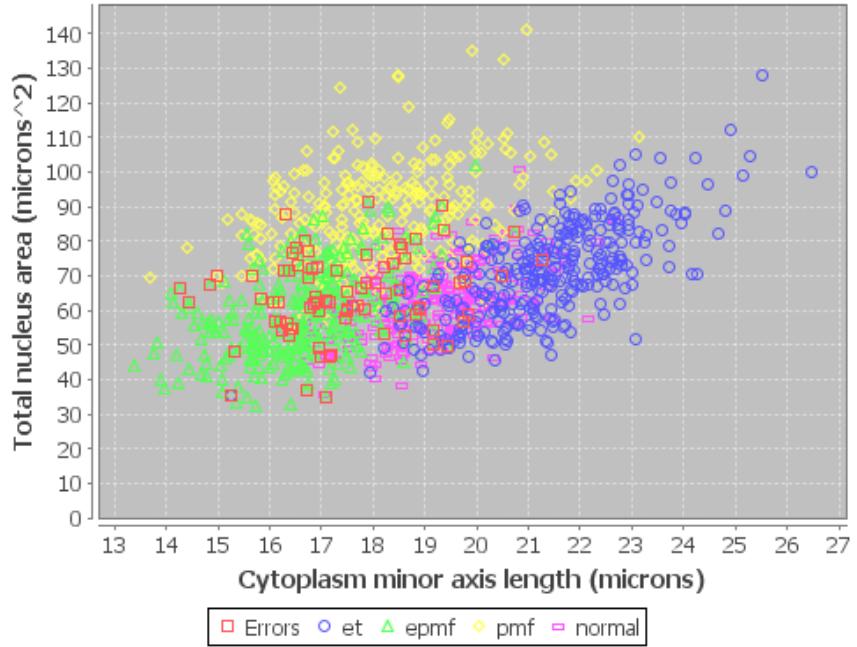


Figure 4.5: A comparison of the Elliptic Fourier feature at the individual megakaryocyte level and after being aggregated into a population feature. Despite individual variations, taking the mean of many megakaryocytes has diluted the feature's usefulness.

(a) A normal megakaryocyte nucleus (b) A PMF megakaryocyte nucleus (c) The aggregated normal nuclei across a TMA (d) The aggregated PMF nuclei across a TMA



(a) Individual megakaryocyte features



(b) Aggregate megakaryocyte features

Figure 4.6: A graph of two important classification features with incorrect predictions shown in red.

4.5.3 Clustering

Due to low sample size, the clustering information collected was not used as a feature for classification since there were not enough samples to produce accurate estimates of their usefulness in classification. By applying the Silhouette Coefficient metric it was possible to determine the best clustering algorithm as affinity propagation with a Silhouette Coefficient of 0.5. This was not unexpected since affinity propagation is the most recently developed of the different algorithms tested.

All algorithms had Silhouette Coefficients of 0.4-0.5, implying that the clusters were relatively distinct. The average intra-cluster distance mostly followed the expectations from the literature, with normal bone marrow having very loose clustering compared to the MPN bone marrow, ET having loose clustering and PMF having the tightest clusters. However, the pre-fibrotic PMF clusters were quite loose compared to all other marrow types. Further investigation is required with higher cluster sample size and verified diagnoses to determine whether this is an outlier or a trend across clusters in pre-fibrotic bone marrow.

CHAPTER 5

Conclusions

As outlined in the introduction, the project's aim was to create a segmentation and analysis algorithm that was quantitative, reliable, robust and practical, and to characterise the clustering and population features of the analysed bone marrow. Good classification results were achieved when classifying megakaryocyte populations, validating the segmentation and feature extraction algorithms. Analysis of clustering and megakaryocyte features can be found in Section 4.5.

5.1 Evaluation of Aims

5.1.1 Quantitative

The classification algorithm was as quantitative as possible, producing numerical estimates of features such as megakaryocyte shape, area and perimeter. Clustering algorithms were specifically chosen that could scale to arbitrary numbers of clusters while still providing an estimate of tightness in the cluster variance metric.

5.1.2 Reliable

The final set of classifiers were quite reliable for megakaryocyte populations, achieving a maximum of 93.4% accuracy in the multilayer perceptron algorithm. The most important statistics for differentiation were found to be the nucleus to cell area ratio and the shape of the best-fitting ellipse around the megakaryocyte.

There were issues obtaining enough samples of megakaryocyte clusters, mainly due to the semi-automated segmentation process. The data obtained suggests that clusters differ between classes in line with the literature, but without enough samples it is difficult to suggest typical values for tightness or number of clusters per patient and class with high confidence.

5.1.3 Robust

The developed segmentation algorithms were robust to variations in the bone marrow images in that mature megakaryocytes were segmented very accurately. However, due to the presence of overlapping and abnormal megakaryocytes in the images it was not possible to develop a sufficiently robust automated algorithm. It proved too difficult to accurately say whether a certain megakaryocyte was overlapping or not, and the requirement of segmenting every megakaryocyte in a patient for clustering purposes meant that they could not simply be ignored either way.

Instead, a mixture of automatic and manual segmentation was relied on that achieved high segmentation accuracy at the cost of time-consuming elimination of segmentation errors. This thesis highlights the areas where further investigation is needed to produce a truly automatic and robust algorithm.

5.1.4 Practical

The features that were extracted from the bone marrow images were similar to those obtained in the literature in order to be as closely aligned as possible with current pathology methods to help pathologists understand and interpret the results. The decision tree classification algorithm can produce a detailed decision tree that pathologists can use to obtain 85% classification accuracy. For these reasons, the developed algorithm is considered practical.

5.2 Future Work

5.2.1 Dataset Improvements

The extracted feature vectors from this project were limited in a number of ways by the dataset used.

As seen in Section 4.5.1, segmentation suffered as a result of segmentation errors. Many of the errors were amplified by the relatively low resolution TMA images used in the dataset. Increasing the optical magnification from 20x to 200-400x would improve cytoplasm and nuclei edge detection.

No three dimensional features such as megakaryocyte volume, 3D clustering and complete surface analysis were analysed due to the restrictions of using two dimensional sections. This also meant that only a single plane of megakaryocyte features was visible: it was impossible to determine whether a particular

megakaryocyte had a different shape when considered in three dimensions or simply appeared abnormally shaped due to where it was cut. Megakaryocytes with missing nuclei had to be ignored due to not knowing whether the cytoplasm had simply been cut across in the wrong plane or whether it was an incorrectly stained object.

Some investigation into potential three dimensional datasets occurred during the project. The work by Takaku *et al.* [13] proves that confocal microscopy was possible at high resolutions but their protocol proved difficult to reproduce and finetune for the samples provided by the Translational Cancer Pathology Laboratory.

Reconstructing a three dimensional view of the bone marrow was more promising - during this project the Translational Cancer Pathology Laboratory also cut and H&E stained 200 sequential bone marrow sections from a patient. It was possible to use image registration for the resultant stack in order to obtain a three-dimensional image with 4-micron vertical resolution. This was a relatively time-consuming process due to a manual digital registration process being required, but with sufficient automation it may be feasible to get three dimensional data with existing techniques.

5.2.2 Better Diagnosis Grading

The provided diagnoses from the Translational Cancer Pathology Laboratory [19] did not take into account PMF grading beyond specifying post-ET PMF and PMF. For optimal detection of pre-fibrotic PMF a graded diagnosis according to the WHO criteria [1] is required. Since there are concerns about inter-observer reliability in the literature [6], images taken from patients who are known to have progressed to PMF would be preferred. As explained in Chapter 2, there is no guarantee that the provided diagnoses were accurate for ET versus pre-fibrotic PMF due to controversy surrounding these disease states. Updated WHO diagnosis guidelines were released in 2016 that could alleviate this problem in future work.

For a true gold standard diagnosis, two biopsies would be taken for each patient separated by a number of years to confirm whether the disease has progressed into advanced PMF or not, in which case the first biopsy can be strongly identified as pre-fibrotic PMF or ET respectively.

5.2.3 Additional Feature Extraction

There are potentially additional unknown criteria beyond the WHO criteria, such as megakaryocyte clustering close to the bone, or changes in the cytoplasm. Machine learning can also be used to generate criteria that are not human-interpretable. Some papers in the field, such as Malherbe *et al.* [19] and Malherbe *et al.* [20], are also beginning to test for the presence of various proteins in abnormal megakaryocytes to see if they are useful in MPN diagnosis.

Additionally, the methodology in this project often relied on mean features rather than characterising outliers separately. It would be useful to have a systematic process to measure the occurrence of outliers across multiple bone marrow images from the same patient to avoid simply throwing that information away entirely. For example, there were problems in this project related to averaging the Elliptic Fourier feature across megakaryocyte populations.

5.2.4 Semi-automated Megakaryocyte Segmentation

The cluster detection algorithms in this project suffered from a lack of samples for training the classifiers. This was in large part due to the manual verification process required to segment every megakaryocyte for a patient. Existing algorithms from the literature cannot circumvent the problems of overlapping megakaryocytes and nucleus colour variations. A semi-automated segmentation process with easily verifiable outputs would allow more cluster examples to be found for better statistical analysis and use in classification.

5.3 Summary

In this work a methodology is presented that produces quantitative, reliable, practical and relatively robust classification results that are comparable with previous papers. Additionally, the usefulness of various clustering algorithms is measured and the observed features are characterised. Overall, this thesis highlights the importance and usefulness of quantitative digital analysis in the area of myeloproliferative neoplasm diagnosis. By developing and testing a variety of different clustering and classification techniques this thesis has contributed a basic framework for pathologists and other researchers to use as groundwork for future investigation.

Bibliography

- [1] S. H. Swerdlow, E. Campo, N. L. Harris, E. S. Jaffe, S. A. Pileri, H. Stein, J. Thiele, and J. W. Vardiman, Eds., *WHO classification of tumours of haematopoietic and lymphoid tissues*. France: IARC Press, 2008.
- [2] B. S. Wilkins, W. N. Erber, D. Bareford, G. Buck, K. Wheatley, C. L. East, B. Paul, C. N. Harrison, A. R. Green, and P. J. Campbell, “Bone marrow pathology in essential thrombocythemia: Interobserver reliability and utility for identifying disease subtypes,” *Blood*, vol. 111, no. 1, pp. 60–70, 2008.
- [3] G. J. Titmarsh, A. S. Duncombe, M. F. McMullin, M. O’Rorke, R. Mesa, F. Vocht, S. Horan, L. Fritschi, M. Clarke, and L. A. Anderson, “How common are myeloproliferative neoplasms? a systematic review and meta-analysis,” *American Journal of Hematology*, vol. 89, no. 6, pp. 581–587, 2014.
- [4] T. Barbui, J. Thiele, F. Passamonti, E. Rumi, E. Boveri, M. Ruggeri, F. Rodeghiero, E. S. d’Amore, M. L. Randi, I. Bertozzi, *et al.*, “Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: An international study,” *Journal of Clinical Oncology*, vol. 29, no. 23, pp. 3179–3184, 2011.
- [5] J. Thiele, H. M. Kvasnicka, L. Müllauer, V. Buxhofer-Ausch, B. Gisslinger, and H. Gisslinger, “Essential thrombocythemia versus early primary myelofibrosis: A multicenter study to validate the who classification,” *Blood*, vol. 117, no. 21, pp. 5710–5718, 2011.
- [6] T. Barbui, J. Thiele, A. Vannucchi, and A. Tefferi, “Problems and pitfalls regarding who-defined diagnosis of early/prefibrotic primary myelofibrosis versus essential thrombocythemia,” *Leukemia*, vol. 27, no. 10, pp. 1953–1958, 2013.
- [7] A. Tefferi, J. Thiele, A. Orazi, H. M. Kvasnicka, T. Barbui, C. A. Hanson, G. Barosi, S. Verstovsek, G. Birgegard, R. Mesa, *et al.*, “Proposals and rationale for revision of the world health organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: Recommendations from an ad hoc international expert panel,” *Blood*, vol. 110, no. 4, pp. 1092–1097, 2007.

- [8] H. M. Kvasnicka and J. Thiele, “Prodromal myeloproliferative neoplasms: The 2008 WHO classification,” *American journal of hematology*, vol. 85, no. 1, pp. 62–69, 2010.
- [9] Y. Wang, R. Xu, G. Luo, and J. Wu, “Three-dimensional reconstruction of light microscopy image sections: Present and future,” *Frontiers of medicine*, vol. 9, no. 1, pp. 30–45, 2015.
- [10] G. S. Travlos, “Normal structure, function, and histology of the bone marrow,” *Toxicologic pathology*, vol. 34, no. 5, pp. 548–565, 2006.
- [11] W. J. Reagan, A. Irizarry-Rovira, F. Poitout-Belissent, A. P. Bolliger, S. K. Ramaiah, G. Travlos, D. Walker, D. Bounous, and G. Walter, “Best practices for evaluation of bone marrow in nonclinical toxicity studies,” *Veterinary Clinical Pathology*, vol. 40, no. 2, pp. 119–134, 2011.
- [12] L. Depalma, “The effect of decalcification and choice of fixative on histiocytic iron in bone marrow core biopsies,” *Biotechnic & histochemistry*, vol. 71, no. 2, pp. 57–60, 1996.
- [13] T. Takaku, D. Malide, J. Chen, R. T. Calado, S. Kajigaya, and N. S. Young, “Hematopoiesis in 3 dimensions: Human and murine bone marrow architecture visualized by confocal microscopy,” *Blood*, vol. 116, no. 15, e41–e55, 2010.
- [14] ——, “Hematopoiesis in 3 dimensions: Human and murine bone marrow architecture visualized by confocal microscopy,” *Blood*, vol. 116, no. 15, e41–e55, 2010.
- [15] B. Ballarò, A. M. Florena, V. Franco, D. Tegolo, C. Tripodo, and C. Valenti, “An automated image analysis methodology for classifying megakaryocytes in chronic myeloproliferative disorders,” *Medical image analysis*, vol. 12, no. 6, pp. 703–712, 2008.
- [16] T.-H. Song, V. Sanchez, H. EIDaly, and N. M. Rajpoot, “A circumscribing active contour model for delineation of nuclei and membranes of megakaryocytes in bone marrow trephine biopsy images,” in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2015, 94200T–94200T.
- [17] M. D. Zarella, D. E. Breen, A. Plagov, and F. U. Garcia, “An optimized color transformation for the analysis of digital images of hematoxylin & eosin stained slides,” *Journal of pathology informatics*, vol. 6, 2015.
- [18] M. Acar, K. S. Kocherlakota, M. M. Murphy, J. G. Peyer, H. Oguro, C. N. Inra, C. Jaiyeola, Z. Zhao, K. Luby-Phelps, S. J. Morrison, *et al.*, “Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal,” *Nature*, vol. 526, no. 7571, pp. 126–130, 2015.

- [19] J. A. Malherbe, K. A. Fuller, A. Arshad, J. Nangalia, G. Romeo, S. L. Hall, K. S. Meehan, B. Guo, R. Howman, and W. N. Erber, “Megakaryocytic hyperplasia in myeloproliferative neoplasms is driven by disordered proliferative, apoptotic and epigenetic mechanisms,” *Journal of clinical pathology*, jclinpath–2015, 2015.
- [20] J. A. Malherbe, K. A. Fuller, B. Mirzai, S. Kavanagh, C.-C. So, H.-W. Ip, B. B. Guo, C. Forsyth, R. Howman, and W. N. Erber, “Dysregulation of the intrinsic apoptotic pathway mediates megakaryocytic hyperplasia in myeloproliferative neoplasms,” *Journal of clinical pathology*, jclinpath–2016, 2016.
- [21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” *ArXiv preprint arXiv:1408.5093*, 2014.
- [22] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [23] I. Arganda-Carreras, V. Kaynig, C. Rueden, J. Schindelin, A. Cardona, and H. S. Seung, *Trainable_segmentation : Releasev3.1.2*, Aug. 2016. DOI: 10.5281/zenodo.59290. [Online]. Available: <https://doi.org/10.5281/zenodo.59290>.
- [24] T. K. Ho, “Random decision forests,” in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, IEEE, vol. 1, 1995, pp. 278–282.
- [25] F. P. Kuhl and C. R. Giardina, “Elliptic fourier features of a closed contour,” *Computer graphics and image processing*, vol. 18, no. 3, pp. 236–258, 1982.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Pas-
sos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [27] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
- [28] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise..”
- [29] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

- [30] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: An update,” *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [31] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.
- [32] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to platt’s smo algorithm for svm classifier design,” *Neural Computation*, vol. 13, no. 3, pp. 637–649, 2001.
- [33] J. R. Quinlan, *C4. 5: Programs for machine learning*. Elsevier, 2014.
- [34] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.
- [35] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

APPENDIX A

Glossary

Biopsy Used as a shorthand for trephine biopsy in this thesis. A trephine biopsy is a core of solid bone marrow extracted with a needle.

Bone marrow Tissue inside bones, broken into ‘yellow’ marrow containing mostly fat cells and ‘red’ bone marrow, which produces the body’s red blood cells. The proportion of yellow to red marrow increases with age. Can be extracted for imaging by taking biopsies.

CALR See **clonal marker**.

CD61 stain A staining method that results in only megakaryocyte cytoplasm being stained red as opposed to the more general hematoxylin and eosin stain. ‘CD61’ refers to an antigen which binds to specific epitopes (antigen binding sites) found on megakaryocyte cytoplasm, allowing it to be stained differently to other cells.

Clonal marker A defective gene that can be used as a disease indicator for certain diseases. For example, most patients with myeloproliferative neoplasms have mutations in the genes that encode the molecules CALR, MPL or JAK2, positively identifying patients with these mutations as candidates for MPNs.

Cytoplasm The biological material inside a cell excluding the nucleus. Although there are different components within the cytoplasm they are not usually distinguishable from an optical microscope.

Differential diagnosis Comparing diagnosis criteria for multiple diseases with similar symptoms in order to find the correct diagnosis.

Eosinophils A subset of white blood cells that are stained well by eosin.

Essential thrombocythosis (ET) A myeloproliferative neoplasm that causes increased megakaryocyte production and thus platelet production, loosely clustered megakaryocytes and staghorn-like megakaryocyte nuclei.

Fibrosis The formation of excess fibrous tissue in the bone marrow. In primary myelofibrosis, this manifests as an increase in reticulin fibers followed by increased

collagen bundles.

Fixative A chemical used in histology to preserve and protect biological tissue from decay, typically for use in slides.

Granulocytes White blood cells that contain granules in their cytoplasm that are visible under optical microscopy.

Hematoxylin & Eosin (H&E) stain The most common tissue staining method for bone marrow. It colours all cell nuclei in blue/purple and cytoplasm in red.

Histology The study of the microscopic anatomy of cells in tissue.

Hyperplasia Refers to an increase in the amount of tissue for a part of the human body, typically in response to a disease. For example, bone marrow hyperplasia means that overall the marrow displays increased numbers of cells such as granulocytes, megakaryocytes and leukocytes and potentially increased average cell size.

Janus kinase 2 (JAK2) mutation See **clonal marker**.

Megakaryocytes A relatively large (50-100 micron) cell type within the bone marrow responsible for platelet production. Contains a large multi-lobulated nucleus. After the nucleus matures through cell division the cytoplasm begins to break off into ribbons which emit platelets into the bloodstream until the cytoplasm is expended.

MPL See **clonal marker**.

Myeloproliferative neoplasm (MPN) A category of diseases that result in increased production of blood cells in the bone marrow, such as platelets and granulocytes.

Pathologist A medical professional in pathology, the study of disease causes. Histopathologists are pathologists who specifically study tissue samples to diagnose disease.

Primary myelofibrosis (PMF) A myeloproliferative neoplasm which causes increased megakaryocyte production, densely clustered megakaryocytes and bulbous megakaryocyte nuclei. Results in increased bone marrow fibrosis as it progresses. Has worse prognosis than essential thrombocythemia. ‘Primary’ refers to the myelofibrosis originating first in the bone marrow as opposed to being a secondary symptom of another disease.

Prognosis The projected outcome of a disease. Usually measured in terms of survival rate, or the proportion of patients alive after a certain number of years.

Tissue microarray (TMA) An imaging technique where multiple smaller biopsies are stained and imaged in the same slide. Allows quicker surveying of the

bone marrow.

APPENDIX B

Various Differential Diagnosis Criteria for Essential Thrombocythosis and Primary Myelofibrosis

B.1 World Health Organisation Criteria

The most widely known criteria are the 2008 WHO criteria for ET and PMF [1]. These criteria are quite broad and describe megakaryocyte features qualitatively instead of quantitatively.

WHO primary myelofibrosis criteria [1, Table 2.04]

Major criteria

1. Presence of megakaryocyte proliferation and atypia,* usually accompanied by either reticulin and/or collagen fibrosis or an increased bone marrow cellularity characterized by granulocytic proliferation and often decreased erythropoiesis (ie, prefibrotic cellular-phase disease)
 2. Not meeting WHO criteria for PV, CML, MDS, or other myeloid neoplasm
 3. Demonstration of JAK2 mutation or other clonal marker, or in the absence of a clonal marker, no evidence of bone marrow fibrosis due to underlying inflammatory or other neoplastic diseases
- * Small to large megakaryocytes with an aberrant nuclear/cytoplasmic ratio and hyperchromatic, bulbous, or irregularly folded nuclei and dense clustering.

Minor criteria

1. Leukoerythroblastosis
2. Increase in serum lactate dehydrogenase level
3. Anemia
4. Palpable splenomegaly

Diagnosis requires meeting all 3 major criteria and 2 minor criteria.

Table B.1: The World Health Organisation diagnosis criteria for primary myelofibrosis, adapted from Swerdlow *et al.* [1]

WHO essential thrombocytosis criteria [1, Table 2.06]

1. Sustained platelet count $\geq 450 * 10^9/L$ during the work-up period
 2. Bone marrow biopsy specimen showing proliferation mainly of the megakaryocytic lineage with increased numbers of enlarged, mature megakaryocytes; no significant increase or left-shift of neutrophil granulopoiesis or erythropoiesis
 3. Not meeting WHO criteria for PV, PMF, CML, MDS, or other myeloid neoplasms
 4. Demonstration of JAK2 mutation or other clonal marker, or in the absence of a clonal marker, no evidence for reactive thrombocytosis
- Diagnosis requires meeting all 4 criteria.
-

Table B.2: The World Health Organisation diagnosis criteria for essential thrombocytosis, adapted from Swerdlow *et al.* [1]

B.2 Alternative Criteria

Thiele *et al.* [5] developed an alternative set of semi-quantitative criteria to try and validate the WHO criteria for ET and pre-fibrotic PMF. Inter-observer reliability was measured at 78% across 230 patients using these criteria [5].

Feature*	ET, %	Early PMF, %
Increased cellularity (age-matched)	10-20	80-100
Neutrophil granulopoiesis		
Increased quantity	≤ 10	50-80
Left-shifting	≤ 10	20-50
Erythropoiesis		
Increased quantity	≤ 10	≤ 10
Left-shifting	≤ 10	10-20
Megakaryopoiesis		
Increased quantity	80-100	50-80
Size		
Small	0	20-50
Medium	10-20	10-20
Large	20-50	20-50
Giant	20-50	10-20
Histotopography		
Endosteal translocation	10-20	20-50
Cluster formation: size		
Small (≥ 3)	10-20	50-80
Large (> 7)	0	20-50
Cluster formation: quality		
Dense	0	20-50
Loose	20-50	50-80
Nuclear features		
Hypolobulation (bulbous/cloud-like)	≤ 10	50-80
Hyperlobulation (staghorn-like)	50-80	≤ 10
Maturation defects	0	50-80
Naked nuclei	20-50	50-80
Fibers		
Increased reticulin (minor fibrosis, grade 1)	0	20-50
Increased collagen	0	0
Osteosclerosis	0	0

* Semiquantitative evaluation (relative incidence, %): 0 indicates usually absent; ≤ 10, rare; 10-20, slight; 20-50, moderate; 50-80, manifest; and 80-100, overt.

Table B.3: Semi-quantitative diagnosis criteria for ET and PMF, adapted from Thiele *et al.* [5]

APPENDIX C

Statistical Measures

C.1 Dice Similarity Coefficient

The Dice similarity coefficient is a measure of set similarity typically used in image segmentation algorithms to estimate their accuracy. It ranges from 0 to 1, with 1 being a perfect segmentation with respect to the test data. It is calculated with the below formula, with X and Y being the algorithm's segmentation and ground truth segmentation respectively.

$$\text{Dice similarity coefficient} = \frac{2|X \cap Y|}{|X| + |Y|}$$

C.2 Inter-observer Reliability

Inter-observer reliability measures how often multiple observers agree with each other on some test. In this thesis it refers to how much pathologists agree with each other on a diagnosis. Inter-observer reliability is usually given as a percentage, with 100% being perfect agreement and 0% being total disagreement. Measures for inter-observer reliability can differ significantly: one common measure is Cohen's κ .

Cohen's κ attempts to correct for the probability that two observers agreed merely by chance. If $\kappa = 1$ then the observers are in complete agreement, and if $\kappa <= 0$ then no agreement has occurred apart from that which could have occurred by chance. The formula is as follows, where p_o is the agreement and p_e is the probability that the agreement occurred by chance:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

C.3 F-Score

A measure of test accuracy that considers two components: the precision, or the number of true positives divided by the total number of true and false positives, and the recall, which is the number of true positives divided by the number of true positives and false negatives. The F-score does not take into account true negatives. It is scaled from 0 to 1, with 1 being perfect precision and recall. The formula is as follows:

$$\text{F-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

C.4 Root Mean Squared Error

The root mean squared error (RMSE) describes the difference between a measured statistic and the observed statistic using the square root of the variance. The formula is as follows, with S being some statistic distribution:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{S}_i - S_i)^2}$$

C.5 Sensitivity and Specificity

A simple statistical measure of the performance of a binary classifier often used in medical diagnosis. The test results are broken up into true positives, false positives, true negatives and false negatives. Sensitivity measures the ability of the test to find true positives while specificity measures the ability to find true negatives as opposed to false diagnoses.

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

Original Honours Proposal

Harry Smallbone

Supervisors: Professor David Smith & Adjunct Professor Bruce Gardiner

1 Background

Proper platelet (thrombocyte) production is essential to the functioning of the human body. Platelets circulate within the bloodstream. If the vascular network is damaged, platelets first attach themselves to the damaged areas of the endothelium, or blood vessel lining. They then release molecules which promote the creation of blood clots around the platelets, preventing further bleeding [5]. Normally, blood clot formation is the first step in healing of damaged tissue. Abnormal platelet numbers in the blood can have serious consequences. If the platelet count falls below about 1/10th of normal levels, it is life threatening due to excessive (or unstoppable) bleeding. It is also potentially life threatening if thrombocytes are more than 3 times normal levels due to the risk of blood vessels being blocked by undesirable blood clots that form with little provocation, such as in myocardial infarction. It is apparent that regulation of platelet numbers in the blood is of critical importance.

Platelets are formed inside the bone marrow by cells called megakaryocytes, which are constantly being formed via cell division. The megakaryocytes mature and then extend 10-20 proplatelet shafts into blood vessels in the marrow (known as sinusoids), which release thousands of platelets into the bloodstream. The megakaryocyte is then spent and its nucleus is released into the bloodstream to be consumed by macrophages [4, 6]. This project is concerned with detecting a group of diseases that affect megakaryocyte production in the bone marrow called myeloproliferative neoplasms (MPNs), typically resulting in an abnormal amount of platelets and associated health risks.

MPNs describe a category of diseases affecting myeloid or bone marrow cell production, of which the most studied are polycythemia vera, essential thrombocyotosis, primary myelofibrosis and myeloid leukemia [8]. Well-known criteria exist for classifying between the various MPNs at later stages of development, such as the identification of the specific JAK2 mutation and high hemoglobin levels for detecting polycythemia vera [8] or a spike in platelet production followed by a drop for primary myelofibrosis as fibrous material begins to inhibit platelet production [9]. An examination of the bone marrow and cellular morphology is necessary in most cases to differentiate essential thrombocyotosis [9]. However, differential diagnosis of early stage or prefibrotic primary myelofibrosis and essential thrombocyotosis is difficult, as both diseases cause an initial spike in platelet production and similar

symptoms: the primary difference between them comes later when primary myelofibrosis begins to create fibrotic tissue [9]. It is important to distinguish between the two as outlined in [8], since treatment strategies are different between the two diseases and primary myelofibrosis has a much lower life expectancy, reported at just 9 years compared to 14 years for essential thrombocytosis in one study [10].

The current criteria for differential diagnosis are provided by the World Health Organization (WHO) [3] and mainly center around feature detection on samples of the megakaryocytes, such as how clustered they are, their shape, and their maturation stage. Using these criteria is subject to considerable false negatives and false positives. In one study of 297 patients, only 78% of diagnoses could be agreed upon by a group of pathologists [1], while another found only 53% agreed when applying the WHO criteria [2]. Another study of 168 essential thrombocytosis patients found that nearly 70% of the patients had either questionable or false positive diagnoses, and a corresponding decrease in their life expectancy as a result [10]. The analysis is further complicated by the lack of cross-correlation values for the WHO criteria.

The chief difficulty in differential diagnosis of these two diseases appears to come from the qualitative criteria from the WHO in early stages of primary myelofibrosis and essential thrombocytosis, such as ‘increased numbers of enlarged, mature megakaryocytes’ [3] or ‘hyperchromatic, bulbous, or irregularly folded nuclei and dense clustering’ instead of measurable quantitative metrics. This may be an issue with the qualitative nature of the current criteria, or the way that they are being applied, but it is clear that quantitation of the criteria would be a step forward. The morphology of the megakaryocytes is markedly different between the two MPNs, with essential thrombocytosis having staghorn-shaped megakaryocytes and primary myelofibrosis having cloud-shaped megakaryocytes. However, using these criteria has not been significantly more successful over the WHO criteria [2]. Currently the most effective metric for diagnosis appears to be categorisation of clusters of megakaryocytes based on whether they are loosely or densely clustered and how many are in each cluster [2].

The WHO criteria have also traditionally been restricted to two-dimensional analysis of slides due to difficulties in imaging the thick bone marrow samples and large megakaryocytes. The thick bone present in the bone marrow is usually removed via a decalcification process using an acidic preservation mixture - however, this alters the morphology of the cells to varying degrees, which is important to differential diagnosis [7]. Three-dimensional imaging is able to access the complete morphology of the megakaryocyte, so decalcification is detrimental to the analysis process.

To enable three dimensional analysis of the bone marrow samples, this project aims specifically to use the technique of confocal microscopy, which utilises a robotic microscope to create optical sections of a sample. This is achieved by treating the sample with a fluorescing agent which reacts to light with specific wavelength emissions. These wavelength emissions are captured at a specific focal plane using a mobile confocal pinhole (aperture) and mirrors, which are moved by the microscope to create a z-stack of images. Contrary to physical sectioning, this technique provides precise horizontal alignment of the captured images and allows for a less decalcified, thicker sample to be used.

Previous work in three dimensional image analysis of the bone marrow has focused on hematopoiesis, or stem cell production, using commercial image analysis software [11, 12]. Typically, software such as ImarisXT is used for its generic image analysis toolbox in thresholding, distance measuring and surface analysis. Despite the high levels of disagreement in differential diagnosis of essential thrombocythemia and primary myelofibrosis [2, 10], a gap exists in the literature on leveraging image analysis to create a better set of criteria than the WHO criteria.

2 Aim

The WHO criteria [3] are currently quite divisive in describing the morphology and clustering patterns of megakaryocytes in affected patients, as shown by poor interobserver reliability measured using interobserver modelling statistics such as pairwise modelling or Cohen κ score. This is due to varying emphases being placed by different pathologists on features such as megakaryocyte morphology versus types of clusters [2]. This seems to be primarily because the criteria lack specific definitions of the relative importance of each criterion, leading each pathologist to apply a different ‘pattern recognition’ scheme to their patients [2].

The aims of this project are as follows:

1. To collect reasonable quality images from the confocal microscopy technique, which is made more difficult by the limitations on sample size for the technique and the need to obtain known good diagnoses when high levels of disagreement and false diagnoses exist. Other considerations for this part of the project will be which fluorophores are most effective for imaging and whether multiple cell detection is possible using multiple fluorophores simultaneously.
2. To analyse the collected images in order to try and move from the qualitative criteria set by the WHO [3] to a quantitative set of metrics that can be used for differential diagnosis.
3. To determine whether the proposed criteria are applicable and useful for the diagnostic process.
4. To attempt to validate the criteria developed in the project on real samples and compare results to previous diagnoses. This will involve collecting ground truths from known diagnoses which can be determined in some cases, as shown by Thiele et al. [?]. The work in [11, 12] suggests that confocal microscopy is a reasonable and novel technique that has not yet been applied to this type of diagnosis.

The data collected for this method is also applicable in other areas, for example examining the megakaryocyte morphology and lifecycle in three dimensions. Creating a model reconstruction pipeline may also be useful for researchers using a commercial package, as these packages frequently have a plug-in environment for reusable work.

3 Method

3.1 Tasks

1. *Image collection [4 weeks]*

Image collection will be done using the confocal microscope available at the Centre for Microscopy Characterisation and Analysis (CMCA). Bone marrow samples will be prepared and provided by the Translational Cancer Pathology Laboratory (TCPL) for imaging purposes. The main subtasks to be completed in this portion of the project are as follows:

- Complete training on use of confocal microscope [1 week].
- Test the confocal microscopy method on control samples to ensure that accurate images can be obtained.
- Collect data using a control, and samples from patients with essential thrombocythosis and primary myelofibrosis.

2. *Development of quantitative criteria [8 weeks]*

- Investigation into what metrics can be measured using available commercial software, and which will need to be extended [3 weeks].
- Preliminary analysis of the collected images to find patterns in the data [3 weeks].
- Development of preliminary quantitative criteria [4 weeks].

3. *Validation of criteria [13 weeks]* Image analysis software will be developed as the main bulk of this section in order to validate the criteria that can detect the different kinds of MPN in the samples. The results of the program will have to be compared to actual diagnoses to confirm a better result than previously found in the literature.

- Investigation into previous data on false positive and false negative rates [2 weeks]
- Development of three-dimensional image analysis algorithms in order to provide a comparison of effectiveness on the data [5 weeks].
- Collection of ground truth data using confirmed diagnoses from the CMCA [4 weeks].
- Investigation of the effectiveness of the image analysis algorithms as compared to existing techniques [2 weeks].

4. *Creation of differential diagnosis pipeline [6 weeks]*

- Modify program to estimate confidence values and add any new criteria that were found to be better for diagnosis [6 weeks].

4 Software and Hardware Requirements

The project may require access to commercial medical image analysis software such as ImarisXT or Fiji, which is anticipated to be provided by the CMCA. For data collection, access to a confocal microscope and bone marrow tissue samples will be required, and is again anticipated to be provided by the CMCA.

References

- [1] Thiele, J, Kvasnicka, HM, Muellauer, L, Buxhofer-Ausch, V, Gisslinger, B & Gisslinger, H 2011, ‘Essential thrombocythemia versus early primary myelofibrosis: a multicenter study to validate the WHO classification’, *Blood*, vol. 117, no. 21, pp. 5710-5718.
- [2] Wilkins, BS, Erber, WN, Bareford, D, Buck, G, Wheatley, K, East, CL, Paul, B, Harrison, CN, Green, AR & Campbell, PJ 2008, ‘Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes’, *Blood*, vol. 111, no. 1, pp. 60-70.
- [3] Swerdlow S, Campo E, Harris N, et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC Press; 2008.
- [4] Machlus, KR & Italiano Jr., JE 2013, ‘The incredible journey: From megakaryocyte development to platelet formation’, *The Journal of Cell Biology*, vol. 201, no. 6, pp. 785-796.
- [5] Born, GVR & Cross, MJ 1963, ‘The Aggregation of Blood Platelets’, *Journal of Physiology*, vol. 168, no. 1, pp. 178-195.
- [6] Patel, SR, Hartwig, JH & Italiano, JE 2005, ‘The biogenesis of platelets from megakaryocyte proplatelets’, *Journal of Clinical Investigation*, vol. 115, no. 12, pp. 3348-3354.
- [7] Naresh, KN, Lampert, I, Hasserjian, R, Lykidis, D, Elderfield, K, Horncastle, D, Smith, N, Murray-Brown, W & Stamp, GW 2006 ‘Optimal processing of bone marrow trephine biopsy: the Hammersmith Protocol’, *Journal of Clinical Pathology*, vol. 59, no. 9, pp. 903-911.
- [8] Tefferi, A & Vainchenker, W 2011, ‘Myeloproliferative Neoplasms: Molecular Pathophysiology, Essential Clinical Understanding, and Treatment Strategies’, *Journal of Clinical Oncology*, vol. 29, no. 5, pp. 573-582.
- [9] Tefferi, A, Thiele, J, Orazi, A, Kvasnicka, HM, Barbui, T, Hanson, CA, Barosi, G, Verstovsek, S, Birgegard, G, Mesa, R, Reilly, JT, Gisslinger, H, Vannucchi, AM, Cervantes, F, Finazzi, G, Hoffman, R, Gilliland, DG, Bloomfield, CD & Vardiman, JW 2007, ‘Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel’, *Blood*, vol. 110, no. 1, pp. 1092-1097.

- [10] Thiele, J, Kvasnicka, HM, Zankovich, R & Diehl, V 2000 ‘Relevance of bone marrow features in the differential diagnosis between essential thrombocythemia and early stage idiopathic myelofibrosis’, *Haematologica*, vol. 85, no. 1, pp. 1126-1134.
- [11] Takaku, T, Malide, D, Chen, J, Calado, RT, Kajigaya, S & Young, N 2010, ‘Hematopoiesis in 3 dimensions: human and murine bone marrow architecture visualized by confocal microscopy’, *Blood*, vol. 116, no. 15, pp. e41-55.
- [12] Acar, M, Kocherlakota, KS, Murphy, MM, Peyer, JG, Oguro, H, Inra, CN, Jaiyeloa, C, Zhao, Z, Luby-Phelps, K, & Morrison, SJ 2015, ‘Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal’, *Nature*, vol. 526, no. 7571, pp. 126-30.

Corrected First Semester Literature Review

27th May 2016

Supervisors: Professor David Smith & Adjunct Professor Bruce Gardiner

Abstract

Two life-threatening diseases, essential thrombocythaemia and primary myelofibrosis, are currently diagnosed manually using images of megakaryocytes. Manual diagnoses using World Health Organisation criteria on these images have problems with inter-observer reliability. Quality digital images of the bone marrow may be able to automatically detect important features about a patient's megakaryocytes, and potentially classify the disease type. These features include 3D nucleus morphology, increased cellularity, clusters of megakaryocytes, cell size and bone marrow hyperproliferation. This paper reviews the current criteria used to differentiate between these two diseases in the early stages and justifies the use of image analysis techniques to investigate possible improvements in these diagnostic criteria.

1 Introduction

Examination of samples of human bone marrow is key to the early differential diagnosis of essential thrombocythaemia (ET) and primary myelofibrosis (PMF). Both diseases belong to a class of diseases called myeloproliferative neoplasms (MPNs), which means that they cause an excess production of clonal myeloid cells in bone marrow. ET and PMF both affect the production of platelets by megakaryocytes in the bone marrow. ET and PMF are particularly difficult to diagnose in the early stages due to overlapping criteria used to diagnose them. Current World Health Organisation (WHO) guidelines recommend analysing images of bone marrow tissue using a set of criteria including megakaryocyte morphology, clustering and frequency to distinguish between the two diseases. Computerised image analysis techniques may be able to assist and perhaps even replace manual identification

and early detection of the two diseases. There are two stages to this process: image acquisition and image analysis using computer vision techniques.

2 Background

Megakaryocytes are essential to the human body as they produce the body's platelets [1]. Platelets circulate within the bloodstream and attach themselves to the walls surrounding damaged blood vessels. They staunch blood loss by fibrin clot formation, which is the first step towards healing the damaged tissues.

Abnormal platelet counts can be life-threatening. Increased platelet levels can cause spontaneous blood clots which may block blood flow to important parts of the body such as the heart or brain. Decreased platelet counts causes reduced blood clotting around damaged vessels, resulting in excessive bleeding.

It is possible to detect MPNs that affect platelet production at an early stage by examining the morphology, production and clustering of megakaryocytes. An image of a normal megakaryocyte can be seen in Figure 1.

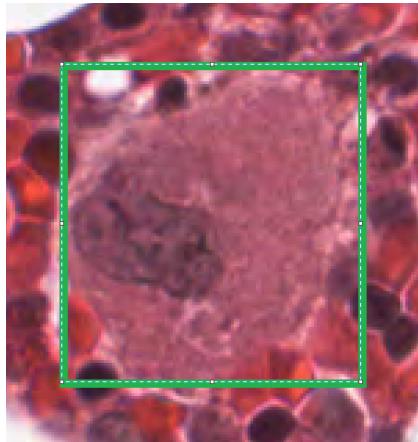


Figure 1: An image of a normal megakaryocyte taken from an H&E stained bone marrow sample. The nucleus is visible inside the surrounding cytoplasm.

Myeloproliferative neoplasms describe a general category of diseases affecting myeloid or bone marrow cell production, of which the most studied are polycythemia vera (PV), essential thrombocythosis, primary myelofibrosis and myeloid leukemia [2]. Myeloid leukemia can be distinguished by checking for the presence of the Philadelphia chromosome, which is missing (i.e.

Philadelphia negative) in the majority of patients with essential thrombocytosis, polycythemia vera, or primary myelofibrosis [3]. In 2008 the World Health Organisation [3] outlined criteria for classifying the remaining three MPNs at various stages of development.

Polycythemia vera is considered the easiest MPN to classify [4]. Virtually all PV patients possess mutations in the Janus 2 kinase (JAK2) gene, which is present in only 50% of other MPNs such as essential thrombocytosis [3]. PV greatly increases erythropoiesis or the production of red blood cells, which provides a major exclusion criteria of elevated haemoglobin levels $> 16.5\text{g/dL}$ [3]. The disease also results in hypercellularity in the bone marrow generally [3] as a minor diagnosis criteria if required. PV eventually progresses to either fibrosis or splenomegaly (enlarged spleen), but can typically be identified at most stages via the major criteria of JAK2 mutation and elevated haemoglobin.

Primary myelofibrosis consists of two major stages: a pre-fibrotic stage characterised by increased platelet levels, and a fibrotic stage with varying levels of reticulin and collagen fibrosis, with progressively decreasing platelet counts. The prognosis ranges from a median survival time of 3-7 years to 10-15 years, depending on whether the disease was diagnosed in the pre-fibrotic or fibrotic stage [3]. In the fibrotic stage it is fairly easy to diagnose PMF due to the presence of excess collagen or reticulin which forms areas of connective tissue across the bone marrow [3]. However, the pre-fibrotic stage is much harder to differentiate due to sharing the indicator of high platelet levels with ET. The JAK2 mutation is present in only 50% of PMF compared to $> 90\%$ of PV cases which limits its use as an exclusion criteria [3]. The WHO recommends looking for megakaryocyte atypia such as dense clustering, bulbous and irregularly folded nuclei and abnormal nucleus/cytoplasm ratios as the major diagnosis criteria [3, Table 2.04].

Essential thrombocytosis is characterised by chronic intermittent periods of elevated platelet levels [3]. In less than 5% of cases ET can progress to fibrotic stages similar to PMF [3]. After diagnosis life expectancy ranges from a median survival time of 10-15 years [3]. The major diagnosis criteria outlined by the WHO are sustained platelet counts $> 450 * 10^9\text{L}$ and the presence of many large 'staghorn-like' megakaryocytes [3]. The JAK2 mutation is present in 40-50% of cases similarly to PMF, so this mutation does not provide any differential diagnostic information [3].

A simplified schematic of the criteria and basic morphology of cells in ET and PMF can be seen in Figure 2, and a comparison of ET megakaryocyte morphology with PMF megakaryocyte morphology can be seen in Figure 3.

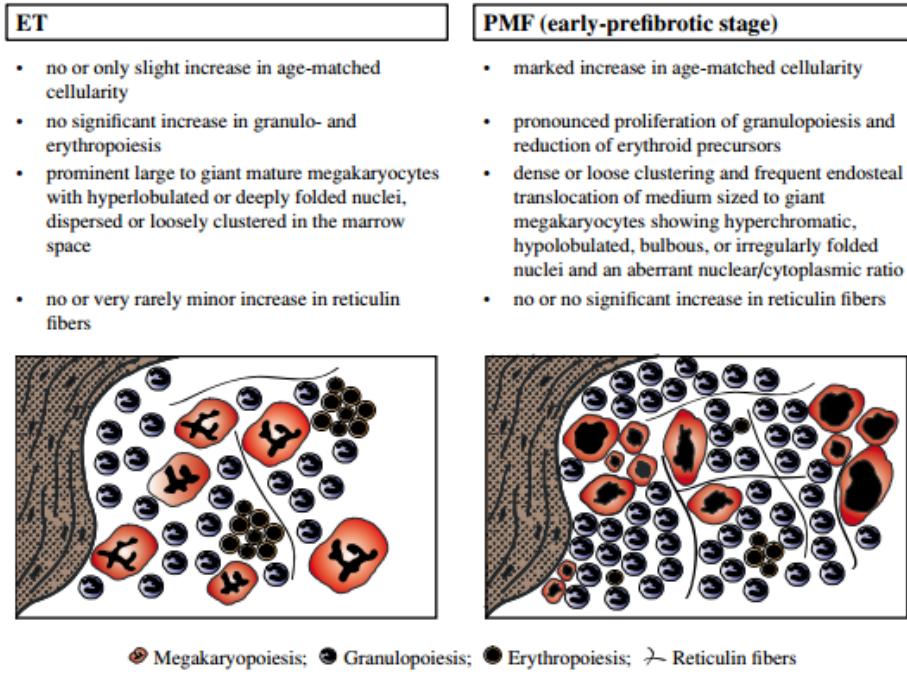


Figure 2: A simplified schematic showing key differences between ET and PMF from Thiele *et al.* [5, Figure 1]

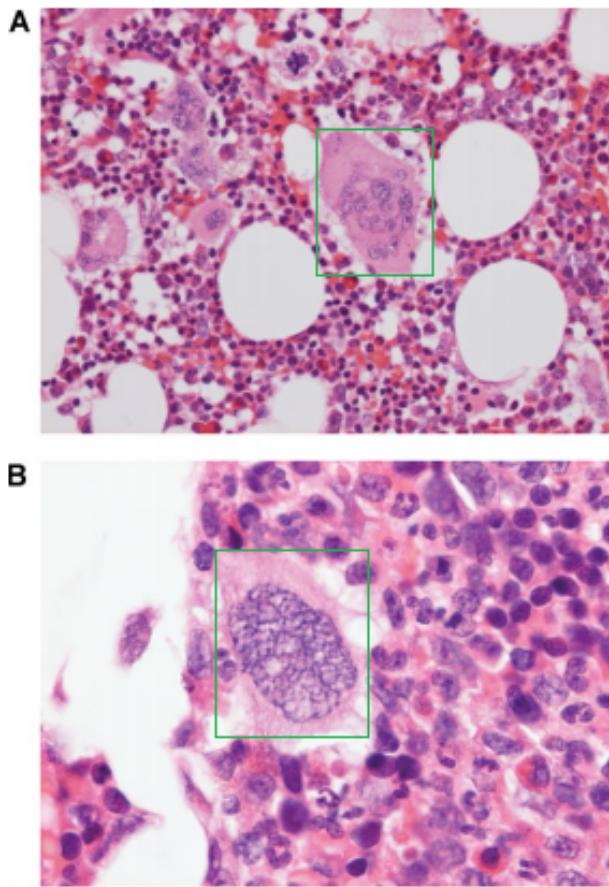


Figure 3: An example staghorn-like megakaryocyte from ET (A) and a cloud-like megakaryocyte from PMF (B) shown inside the green box (modified from an H&E stained bone marrow section taken by Wilkins *et al.* [6])

Clearly, pre-fibrotic PMF and ET require very careful examination of megakaryocyte morphology and megakaryopoiesis for differential diagnosis due to overlapping major criteria.

It is important to distinguish between the two diseases for prognostic and treatment reasons. Barbui *et al.* [7] explain this in more detail using a study of 1,104 ET/PMF cases with bone marrow histology carried out according to the WHO criteria. Firstly, ET has a greater incidence compared to PMF of thrombocytosis or excess platelets and associated complications with blood clots [7]. PMF has a higher rate of leukemic transformation which progresses more rapidly than fibrosis and contributes significantly to deaths post-diagnosis of PMF [7, Figure 1]. The 15-year survival rate of positively diagnosed ET patients is roughly 75% compared to 44% for PMF [7, Table

2]. The differential diagnosis of ET and early pre-fibrotic PMF is key to determining potential risk factors and prognosis for patients.

Previous application of image analysis to this specific problem was performed in Tripodo *et al.* [8] using a hybrid 2D morphologically based image segmentation algorithm. The algorithm applied thresholding and wavelet transforms to accurately identify the cell nucleus and cytoplasm of megakaryocytes [8]. The paper found that it was possible to accurately identify abnormal megakaryocytes with sensitivity and specificity 98.48% and 96.65% respectively using a combination of cell area, cell perimeter, the eccentricity of the cell contour and its complexity, as measured by the difference between the original contour and its approximation by using an elliptic Fourier transform [8]. They achieved sensitivities of 88.2% and 90.2% for detection of ET and PMF respectively using a combination of measures include cell and nuclear area and perimeter, ellipsis eccentricity and cell/nucleus ratio [8]. However, this paper did not perform any investigation into 3D megakaryocyte reconstruction and predates the release of the WHO criteria, so does not apply criteria such as clustering to the results of its segmentation algorithm [8].

3 Review of WHO criteria

The 2008 criteria for differential diagnosis of MPNs are provided by the World Health Organization (WHO) [3] and mainly focus on manual feature detection by trained pathologists on megakaryocyte samples, such as how clustered the megakaryocytes are, their shape, and their maturation stage [4, Table 5]. There is little quantification of these criteria or guidance on how to interpret these criteria. For example, there are no correlation coefficients, conditional probabilities or sample sizes for the morphological criteria. Furthermore, there is no definition of what constitutes a cluster beyond ‘dense’ or ‘loose’ clustering of megakaryocytes [3].

In order to diagnose ET, the WHO criteria state that the patient must demonstrate sustained increased platelet counts, demonstration of a clonal marker such as the JAK2 mutation or the evidence-based exclusion of other diseases, bone marrow showing proliferation of large, mature megakaryocytic cells and cannot meet all of the criteria for another MPN [3]. There are three major criteria for PMF which must all be met for a true positive diagnosis: presence of megakaryocyte proliferation and atypia or fibrosis for late-stage PMF, demonstration of a clonal marker such as the JAK2 mutation or the evidence-based exclusion of other diseases, and not meeting the criteria of other MPNs such as polycythemia vera [3]. The presence of excess fibrotic

tissue can positively identify PMF but this is not present in its early stages [3]. Thus in order to diagnose ET in the case of possible pre-fibrotic PMF the main comparison criteria is megakaryopoeisis and megakaryocyte morphology [3]. These criteria can be seen in Table 1.

WHO primary myelofibrosis criteria [3, Table 2.04]

Major criteria

1. Presence of megakaryocyte proliferation and atypia,* usually accompanied by either reticulin and/or collagen fibrosis, or, in the absence of significant reticulin fibrosis, the megakaryocyte changes must be accompanied by an increased bone marrow cellularity characterized by granulocytic proliferation and often decreased erythropoiesis (ie, prefibrotic cellular-phase disease)
2. Not meeting WHO criteria for PV, CML, MDS, or other myeloid neoplasm
3. Demonstration of JAK2 mutation or other clonal marker, or in the absence of a clonal marker, no evidence of bone marrow fibrosis due to underlying inflammatory or other neoplastic diseases

Minor criteria

1. Leukoerythroblastosis
2. Increase in serum lactate dehydrogenase level
3. Anemia
4. Palpable splenomegaly

Diagnosis requires meeting all 3 major criteria and 2 minor criteria.

* Small to large megakaryocytes with an aberrant nuclear/cytoplasmic ratio and hyperchromatic, bulbous, or irregularly folded nuclei and dense clustering.

WHO essential thrombocythemia criteria [3, Table 2.06]

1. Sustained platelet count $\geq 450 * 10^9/L^*$
2. Bone marrow biopsy specimen showing proliferation mainly of the megakaryocytic lineage with increased numbers of enlarged, mature megakaryocytes; no significant increase or left-shift of neutrophil granulopoiesis or erythropoiesis
3. Not meeting WHO criteria for PV, PMF **, CML, MDS, or other myeloid neoplasm
4. Demonstration of JAK2 mutation or other clonal marker, or in the absence of a clonal marker, no evidence for reactive thrombocytosis

Diagnosis requires meeting all 4 criteria.

* During the work-up period.

** Requires the absence of relevant reticulin fibrosis, collagen fibrosis, peripheral blood leukoerythroblastosis, or markedly hypercellular marrow for age accompanied by megakaryocyte morphology that is typical for PMF- small to large with an aberrant nuclear/cytoplasmic ratio and hyperchromatic, bulbous or irregularly folded nuclei and dense clustering.

Table 1: A table of the criteria for essential thrombocythemia and primary myelofibrosis taken from Swerdlow *et al.* [3]

The literature shows some controversy around these criteria due to variable agreement among practitioners using the criteria for diagnosis. In one study of 297 patients, only 78% of histologic diagnoses using similar standardised criteria could be agreed upon by a group of pathologists [5], while another found only 53% agreed when applying the WHO criteria [6]. Barbui, Thiele, Vannucchi, and Tefferi [9] reviewed multiple studies of observer reliability and found concordance rates ranging from 35% to 63% to as high as 70%. Barbui *et al.* [7] used seven cross-validated centres in order to achieve reasonable confidence in applying the WHO criteria and had a concordance rate of 81%. In general, better outcomes are obtained with pathologist training, pretesting of observers to ensure that interobserver variability is minimised, and by applying strict exclusion criteria to patients that are admitted to the study [9]. The WHO criteria are not yet easily applicable to general practice as shown by these concordance studies.

Some research has been conducted into developing measurable rules and indicator strength for the two diseases. The concrete criteria developed by Thiele *et al.* [5] found a diagnosis consensus rate of 73% across 120 patients. This study had a small sample size and some patients which were difficult to strongly categorise as ET or PMF, but provides insight into potential ranges for diagnosis criteria [5].

Wilkins *et al.* [6] studied interobserver reliability for each individual criteria, using three-point scales for features such as cluster sizes and morphology. This represents the relative importance that each criteria had to a successful diagnosis between ET and PMF. The observers were three hematopathologists with over 10 years experience and specialisations in MPNs. Of the WHO criteria, detecting bare megakaryocyte nucleii and the number and size of megakaryocyte clusters was most reliable for diagnosis [6].

However, this study was heavily criticised by Kvasnicka and Thiele [10] for not measuring intraobserver reliability, having low levels on agreement on basic features like erythropoiesis (red blood cell production) and taking smaller (0.5cm) than recommended (1.5cm) biopsies leading to less opportunity to observe clustering and for failing to standardise megakaryocyte features throughout the paper. Kvasnicka and Thiele [10] also discuss followup studies which showed that the original study group had a high proportion (~60%) of late-stage myelofibrosis patients which may have skewed the results. A recent review [9] calls for proposing ‘scientifically sound, objective, repeatable quantitative criteria for the prefibrotic variant of PMF followed by a corresponding prospective clinico-pathological study’ to resolve these issues.

4 Image acquisition review

Many different imaging techniques such as X-ray microtomography, magnetic resonance imaging, optical projection tomography, optical imaging and confocal microscopy exist for image acquisition and analysis of bone marrow. Each technique has varying characteristics, usually a tradeoff between image resolution, accurate isolation of megakaryocytes and original cell morphology preservation.

We restrict imaging techniques considered here according to the following criteria: high cellular resolution (1-10 microns), ability to stain the cytoplasm/nuclei of megakaryocytes, as little tissue morphological alteration as possible and ability to reconstruct 3D images. We prefer techniques which are likely to enter clinical practice from research or that are already in use by practitioners to improve the practicality of any result.

These criteria are justified by the nature of bone marrow and megakaryocyte imaging. Bone marrow exists inside thick optically opaque trabecular bone and has many different functions apart from megakaryocyte production. Bone marrow tissue consists of red marrow, containing erythropoietic (red blood cell producing) and myelopoietic (white blood cell producing) cells and yellow marrow, containing connective tissues and fat cells. The fraction of yellow marrow to red marrow increases with age, and is about 50% in an adult. Identifying a large megakaryocyte population can be difficult because they are relatively large and rare cells within the red marrow and must also be distinguished from the other types of cells.

Megakaryocytes are relatively large cells, ranging from 50-100 microns thick with a diameter of 20-49 microns [11] at maturity, which means that techniques with high resolutions are preferred in order to fully resolve 3D features. For comparison, smaller cells such as neutrophils can range from 7-15 microns in diameter. This favours techniques like optical and confocal microscopy which have 1-4 micron resolution [12]. Generally, we prefer techniques which offer the highest resolution such as 1 micron resolution due to the potential for new diagnostic criteria.

Since the WHO criteria describe clustering as an important factor for differential diagnosis [3], imaging techniques which alter cell position must be excluded. Megakaryocyte morphology should also be preserved as much as possible for important features such as staghorn-like or cloud-like megakaryocytes.

Of current techniques the two methods most suitable according to these criteria are optical imaging with 3D reconstruction of 2D sections of tissue and confocal microscopy [12].

4.1 Optical microscopy

Optical microscopy is a commonly used method to acquire images of bone marrow. Travlos [13] provides an overview of this technique as accepted in modern practice. First, a trephine biopsy of solid bone marrow tissue is taken and placed in a fixative solution such as 10% neutral buffered formalin [13] to harden and preserve the cells inside the marrow. Next, acid is used to decalcify and remove the opaque minerals from bone in the sample, which would otherwise block the microscope light. The sample is then cut into physical sections usually about 3-5 microns thick [13] and placed sequentially onto a series of glass slides to feed into the microscope. Each slide undergoes a clearing and staining process using chemicals such as hematoxylin and eosin (H&E) which bind to the cytoplasm and nucleus of the megakaryocytes, distinguishing them in different colours from the background cells. Finally, the optical microscope scans the slides into a stack of images. The stack of images must then be digitally aligned. An example megakaryocyte slice can be seen in Figure 4.

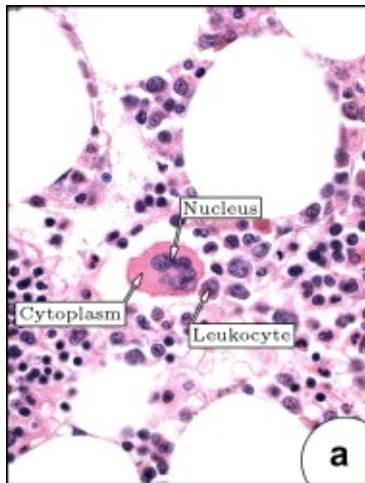


Figure 4: An example megakaryocyte and adjacent leukocyte from an H&E stained bone marrow section taken by Ballarò *et al.* [14]

Travlos [13] discusses some of the disadvantages of this approach. One of the disadvantages mentioned is the need for precise decalcification and fixation protocols to avoid damaging cellular structure and staining ability [13]. Acid treatment is required to optically opaque minerals from the bone but damages protein epitopes in the process, which are binding spots for staining antibodies. Using acid reduces the effectiveness of stains depending on which epitopes are damaged and consequently weakens the image if not managed

correctly. In some cases, the sectioning and staining process may have to be repeated multiple times to test which fixative and acid combination yields the best results [15].

Reagan *et al.* [16] analyse further shortcomings of the optical imaging process. Some artefacts can occur during the sectioning process, although great care by pathologists is taken to try and avoid them. Typically sections are taken using a microtome which can cut physical sections down to 1 micron thick. However, 1 micron sections break apart when using common fixatives such as paraffin so 4 micron sections are the standard. A common approach is to use the frozen section procedure in order to keep the bone marrow hard enough to be cut by the microtome blade. A single section is cut from the larger frozen sample and transferred into water to thaw before being mounted on a slide using wax for storage. These sections are on the order of microns thick and can easily be lost, torn or otherwise damaged during this process [16]. Again, smaller sections exacerbate this problem.

The average mature megakaryocyte is around 50-100 microns thick, and each physical section is typically around 4-5 microns [16]. Hundreds of sections must be taken to fully cover a representative bone marrow sample, and the exact boundaries of the megakaryocytes may be missed. This can lead to problems in creating 3D images since alignment points may be missing. 3D contours are also lacking in fine detail which may hide useful diagnostic features. Using physical sections means that the individual cells may be cut and stained unevenly. The fixative used to preserve the sample must be chosen carefully: widely used fixatives such as formaldehyde can shrink the sample by up to 4% or more [12].

The advantages of this process are mainly in that it has a high resolution (1-5 microns) compared to other techniques like magnetic resonance imaging, and that it clearly separates megakaryocyte nuclei and cytoplasm in different colours. There are a number of standard staining compounds which can be used to colour cell nuclei, cytoplasms, red blood cells, collagen and other interesting cells. For megakaryocyte staining the most common and effective is hematoxylin and eosin (H&E) staining which results in coloured cell nuclei and cytoplasms only. Although optical imaging is able to colour the megakaryocytes differently visually, it still requires a separate digital analysis step to perform object recognition on the cytoplasm and nuclei for further analysis. For the most part it preserves cell position and structure apart from previously described artefacts such as shrinkage, tearing, scratching, distortions and misalignment. Optical imaging with H&E staining also results in staining non-megakaryocyte cells such as granulocytes and lymphocytes which can be a marker for ET and PMF [4].

Optical imaging requires each section to be imaged separately so a 3D

registration process is necessary to create an image stack. Linear transformations such as rotation and translation as well as non-linear transformations may occur during sectioning: Brändle [17] describes common methodology used to resolve this, examined later in Section 5. Assuming that only linear transformations have occurred, key features common to sequential slides such as large cells can be used to reconstruct the original image with high accuracy [17]. However, non-linear artefacts such as sample folding and warping occur frequently, necessitating more in-depth distortion correction models.

4.2 Confocal microscopy

Confocal laser scanning microscopy or confocal microscopy is an alternative imaging technique which allows for large optical sections to be imaged by capturing light emitted at a specific sample depth. There are two primary differences from optical microscopy. Firstly a pinhole camera is added to capture light emitted at a specific plane instead of capturing all light that passes through the sample. Secondly, a laser used to excite fluorescent probes in the sample instead of a lamp. In this approach, a solid sample of tissue is first stained with antibodies that attach fluorescent probes to the megakaryocytes. Inside the microscope the laser is focused on a specific point in the sample using mirrors to excite the probes which emit light into the pinhole. The pinhole eliminates out of focus light and records the captured light using sensitive detectors. This is repeated for each point on the sample until an image is built up. The objective lens is moved to examine different focal planes to create 3D images.

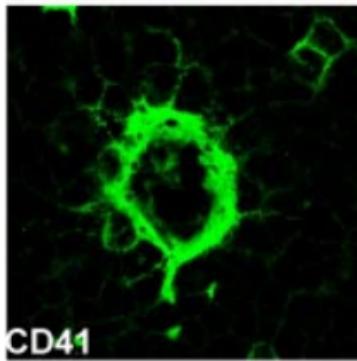


Figure 5: An example megakaryocyte stained with CD41 antibodies and imaged using confocal microscopy by Winter *et al.* [18]

Confocal microscopy results in a stack of well-aligned 3D images. By using staining antibodies, megakaryocytes and their nuclei can be precisely

identified from the background of the images. The optical sectioning resolution is flexible and a volume of interest can be specified to capture different parts of the bone marrow. This process reduces the number of physical samples required compared to optical microscopy as samples can be imaged in 150-300 micron blocks at 1 micron resolution [19] instead of many 1-5 micron sections. All of the megakaryocytes inside a 150-300 micron sample can be imaged *in situ* without the linear and non-linear artefacts introduced by the high number of sections in optical imaging. The sample does not need to be decalcified with acids that damage protein epitopes as the emitted light from the fluorescent probes can pass through the minerals in the bone.

The principal limitation of this technique is the maximum thickness of the sample that emitted light can pass through. Takaku *et al.* [19] achieved sample thicknesses of 150 microns, which limits the potential number of visible megakaryocytes, which are 50-100 microns thick and may be cut off. To mitigate this, samples can be cut 200-300 microns thick and imaged from either side to obtain an overall 3D stack.

Another limitation is the inability to detect other areas of the section which have not been tagged by antibodies. Confocal microscopy is generally limited to a maximum of four different fluorescent frequencies per sample. This is because the fluorescent probes emit at a range of input and output frequencies. Widely used probes can have overlapping frequency ranges and subsequently be difficult to distinguish in the processing stage, commonly referred to as spectral bleed through. As a result, although the resultant image colours can be specifically associated with cell labels like megakaryocytes or granulocytes, only a maximum of four such colours are recommended with most available confocal microscopes. The marked cells and fluorophores used must be chosen with care to maximise potential sources of diagnostic criteria while also labelling megakaryocytes nucleii and cytoplasm separately.

5 Image analysis

Much of the work to date as outlined in Section 3 has focused on optical imaging. As such, according to the criteria in Section 4 optical imaging is the preferred technique despite the advantages of confocal microscopy.

5.1 Image registration

Using optical imaging requires a registration or 3D image alignment step due to the usage of separate sequential sections. As the slides are taken separately there is a high chance of shearing, rotational and translational

errors occurring during image acquisition. There are two major categories of image registration algorithms commonly used in medical imaging: rigid and nonrigid.

5.1.1 Rigid registration

Rigid registration is a class of algorithms that restricts the set of transformations used to align the image to linear transformations such as rotation and translation [17]. Sequential tissue sections are not direct linear transforms as they are slices through different levels of the tissue. Reasonable accuracy can be achieved through the use of key landmarks such as large cells which are common across a sequence of slides and assuming minimal non-linear geometric transformations. Prescott *et al.* [20] used rigid registration to construct 3D models of a human cochlear sample. One advantage of using this technique is that it is sample-independent: as long as key registration points are identified no specific information about possible non-linear transformations is required. This is especially useful when many key features are present such as large blocks of similar colour between sections.

5.1.2 Non-rigid registration

Non-rigid registration algorithms assume some non-linear transformations to correct for geometric distortions such as irregular cell walls across sections, folds in the section and stretching of tissue samples. There are many different models to apply which correct for distortions in the sectioning process and in most cases they are problem-specific. In general, they first segment the images to detect key features, and then apply some domain-specific knowledge to correct distortions across slides. These algorithms typically take more time to evaluate due to the need to analyse the images in more detail than simple rigid registration. They can also be difficult to evaluate visually since the structure of the image is altered. Although it is difficult to compare results from different types of tissues, one paper found a 3D registration accuracy of 82% [21] using a mixture of computed deformation fields, colour matching and feature extraction.

5.2 Segmentation

Image segmentation refers to the process of applying algorithms to transform an input image into segmented regions of pixels which represent areas of interest. Image segmentation and classification is a broad area of current research. This project requires the analysis of the morphology of each

individual megakaryocyte: the ideal image segmentation algorithm is able to distinguish megakaryocyte boundaries from the surrounding background even when cell boundaries overlap as well as differentiate the cell nucleus from its surrounding cytoplasm. For this specific problem, clustering algorithms for segmentation purposes are less useful without preprocessing as the staining process results in high colour variation across cells.

Classical image segmentation techniques have already been applied to the problem of segmenting megakaryocytes. Ballarò *et al.* [14] obtained high quality megakaryocyte classification rates of 98.4% using a hybrid image segmentation algorithm. They used H&E stained bone marrow images, which coloured the nuclei and cytoplasm of megakaryocytes as well as other cells like leukocytes and granulocytes in shades of purple [14]. Thresholding was used on the images converted to grayscale to put the cell nuclei and cytoplasm of the megakaryocytes into two distinct colours and remove extraneous background colours from the images. Next, standard morphological techniques were used to detect edges, increase contrast, isolate and smooth the megakaryocytes. A wavelet function was used to find the cell nucleus edges more precisely. This resulted in clear boundaries for detected megakaryocytes with high accuracy edges for both cytoplasm and nucleus [14]. This technique is non-parametric and does not require a training period. Ballarò *et al.* [14] validated the algorithm using a database of 297 bone marrow microphotographs of 763x577 pixels taken at 400x magnification. The database was specifically selected for a variety of normal, PMF and ET patients with different megakaryocyte morphologies and the algorithm had average recognition rates of 98.4%, 91.7% and 86.2% respectively.

One alternative that has been applied extensively in other fields is to use an artificial neural network with supervised training for segmentation. Although it has not yet been applied to megakaryocytes specifically, it has been used for image segmentation in other areas such as hippocampus and prostate segmentation [22], [23]. The most challenging aspect of this approach is obtaining training data for the neural network to process: either manual labelling or another segmentation technique is required to obtain a large training and testing dataset. However, a gap exists in current research in applying neural networks to megakaryocyte segmentation and with proper testing and validation from pathologists it is a viable method of detection.

6 Conclusion

There has been continued debate on the application of existing WHO criteria to differential diagnosis of essential thrombocytosis and primary myelof-

brosis focused around the key indicators of megakaryocyte clustering and morphology. Studies on inter-observer reliability show that manual identification from bone marrow images can be unreliable without extensive cross-validation. There is little in the way of quantitative criteria, cross-correlation values or automated tooling for this problem. Previous investigation into using computer vision techniques for this area has been limited to pure 2D image segmentation for megakaryocytes without investigating its impact on the WHO criteria. Clearly, a gap exists in the research for an investigation into the impact of using computer vision techniques to segment and analyse megakaryocytes on the WHO criteria. Areas of interest which have either not been explored or need further investigation include detection reliability, cross-correlation values and automatic extraction of key features for medical practitioners and researchers.

References

- [1] K. R. Machlus and J. E. Italiano, “The incredible journey: From megakaryocyte development to platelet formation,” *The Journal of Cell Biology*, vol. 201, no. 6, pp. 785–796, 2013.
- [2] A. Tefferi and W. Vainchenker, “Myeloproliferative neoplasms: Molecular pathophysiology, essential clinical understanding, and treatment strategies,” *Journal of Clinical Oncology*, vol. 29, no. 5, pp. 573–582, 2011.
- [3] S. H. Swerdlow, E. Campo, N. L. Harris, E. S. Jaffe, S. A. Pileri, H. Stein, J. Thiele, and J. W. Vardiman, Eds., *WHO classification of tumours of haematopoietic and lymphoid tissues*. France: IARC Press, 2008.
- [4] A. Tefferi, J. Thiele, A. Orazi, H. M. Kvasnicka, T. Barbui, C. A. Hansson, G. Barosi, S. Verstovsek, G. Birgegard, R. Mesa, *et al.*, “Proposals and rationale for revision of the world health organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: Recommendations from an ad hoc international expert panel,” *Blood*, vol. 110, no. 4, pp. 1092–1097, 2007.
- [5] J. Thiele, H. M. Kvasnicka, L. Müllauer, V. Buxhofer-Ausch, B. Gisslinger, and H. Gisslinger, “Essential thrombocythemia versus early primary myelofibrosis: A multicenter study to validate the who classification,” *Blood*, vol. 117, no. 21, pp. 5710–5718, 2011.

- [6] B. S. Wilkins, W. N. Erber, D. Bareford, G. Buck, K. Wheatley, C. L. East, B. Paul, C. N. Harrison, A. R. Green, and P. J. Campbell, “Bone marrow pathology in essential thrombocythemia: Interobserver reliability and utility for identifying disease subtypes,” *Blood*, vol. 111, no. 1, pp. 60–70, 2008.
- [7] T. Barbui, J. Thiele, F. Passamonti, E. Rumi, E. Boveri, M. Ruggeri, F. Rodeghiero, E. S. d’Amore, M. L. Randi, I. Bertozzi, *et al.*, “Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: An international study,” *Journal of Clinical Oncology*, vol. 29, no. 23, pp. 3179–3184, 2011.
- [8] C. Tripodo, C. Valenti, B. Ballarò, Z. Rudzki, D. Tegolo, V. Di Gesù, A. Florena, and V. Franco, “Megakaryocytic features useful for the diagnosis of myeloproliferative disorders can be obtained by a novel unsupervised software analysis,” 2006.
- [9] T. Barbui, J. Thiele, A. Vannucchi, and A. Tefferi, “Problems and pitfalls regarding who-defined diagnosis of early/prefibrotic primary myelofibrosis versus essential thrombocythemia,” *Leukemia*, vol. 27, no. 10, pp. 1953–1958, 2013.
- [10] H. M. Kvasnicka and J. Thiele, “Prodromal myeloproliferative neoplasms: The 2008 who classification,” *American journal of hematology*, vol. 85, no. 1, pp. 62–69, 2010.
- [11] R. F. Levine, K. C. Hazzard, and J. D. Lamberg, “The significance of megakaryocyte size,” *Blood*, vol. 60, no. 5, pp. 1122–1131, 1982.
- [12] Y. Wang, R. Xu, G. Luo, and J. Wu, “Three-dimensional reconstruction of light microscopy image sections: Present and future,” *Frontiers of medicine*, vol. 9, no. 1, pp. 30–45, 2015.
- [13] G. S. Travlos, “Normal structure, function, and histology of the bone marrow,” *Toxicologic pathology*, vol. 34, no. 5, pp. 548–565, 2006.
- [14] B. Ballarò, A. M. Florena, V. Franco, D. Tegolo, C. Tripodo, and C. Valenti, “An automated image analysis methodology for classifying megakaryocytes in chronic myeloproliferative disorders,” *Medical image analysis*, vol. 12, no. 6, pp. 703–712, 2008.
- [15] L. Depalma, “The effect of decalcification and choice of fixative on histiocytic iron in bone marrow core biopsies,” *Biotechnic & histochemistry*, vol. 71, no. 2, pp. 57–60, 1996.

- [16] W. J. Reagan, A. Irizarry-Rovira, F. Poitout-Belissent, A. P. Bolliger, S. K. Ramaiah, G. Travlos, D. Walker, D. Bounous, and G. Walter, “Best practices for evaluation of bone marrow in nonclinical toxicity studies,” *Veterinary Clinical Pathology*, vol. 40, no. 2, pp. 119–134, 2011.
- [17] K. Brändle, “A new method for aligning histological serial sections for three-dimensional reconstruction,” *Computers and biomedical research*, vol. 22, no. 1, pp. 52–62, 1989.
- [18] O. Winter, K. Moser, E. Mohr, D. Zotos, H. Kaminski, M. Szyska, K. Roth, D. M. Wong, C. Dame, D. M. Tarlinton, *et al.*, “Megakaryocytes constitute a functional component of a plasma cell niche in the bone marrow,” *Blood*, vol. 116, no. 11, pp. 1867–1875, 2010.
- [19] T. Takaku, D. Malide, J. Chen, R. T. Calado, S. Kajigaya, and N. S. Young, “Hematopoiesis in 3 dimensions: Human and murine bone marrow architecture visualized by confocal microscopy,” *Blood*, vol. 116, no. 15, e41–e55, 2010.
- [20] J. Prescott, M. Clary, G. Wiet, T. Pan, and K. Huang, “Automatic registration of large set of microscopic images using high-level features,” in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, IEEE, 2006, pp. 1284–1287.
- [21] C.-W. Wang, E. B. Gosno, and Y.-S. Li, “Fully automatic and robust 3d registration of serial-section microscopic images,” *Scientific reports*, vol. 5, 2015.
- [22] A. Brebisson and G. Montana, “Deep neural networks for anatomical brain segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 20–28.
- [23] S. Liao, Y. Gao, A. Oto, and D. Shen, “Representation learning: A unified deep learning framework for automatic prostate mr segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, Springer, 2013, pp. 254–261.