Computerised Renal Interstitial Fibrosis Quantification System

MONASH University

Tey Wei Keat

Department of Electrical and Computer Systems Engineering
Monash University

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

September, 2018

I dedicate this thesis to my parents, for being the cornerstone to my narrative.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

.....

Tey Wei Keat

September, 2018

Acknowledgements

As I put the finishing touches to this thesis, I would like to reflect on the good people who have offered me support and guidance in my research over this past four years.

Firstly, I would like to express my sincerest appreciation to my main research supervisors, Associate Professor Melanie Po-Leen Ooi and Associate Professor Ye Chow Kuang. They have never failed to steer me onto the right track with my research progress, even during times when I could not see the way forward. Thank you very much for being the parent figures in my PhD journey.

I am also particularly grateful for the kind guidance provided by my co-supervisor, Associate Professor Joon Joon Khoo, in my crash course into renal biology. Even amidst her busy schedule split between her university and hospital duties, she has managed to assist me in collecting the biopsy data and pathologist evaluations vital for this research. I would also like to thank Dr. Vineetha Kalavally for her kind reminders and supervision, especially towards the conclusion of my research and thesis writing.

I wish to acknowledge the pathologists who have agreed to participate in this study: Dr. Subathra Sabaratnam, Dr. Kam Pik Chia, Dr. Ming Yen Lee, Dr. Arati Devi Chhetri, Dr. Nurul Akmar and Dr. Hanizah Khairan. I am indebted to them for their willingness to contribute their very valuable time to complete the multiple rounds of biopsy evaluation.

My special thanks are extended to the administrative staff from the Monash University Malaysia School of Engineering —Ms. Krisnakumari Perumal, Ms. Wan Nurul Rukiah and Dr. Poh Phaik Eong, for their aid in my candidature-related matters. My gratitude is also given to the Malaysian Ministry of Higher Education Fundamental Research Grant Scheme and Monash University Malaysia for funding the research.

To my amazing friends and wonderful colleagues, especially the GuruGuruGuru peeps, thank you for the company and for always reminding me to keep sane.

Finally, warmest thanks to my family for never doubting me, even when I did so myself.

Thank you.

Abstract

Kidney diseases, whether they are caused by genetic predisposition or injury to the kidney, can lead to severe deterioration of human health if left undetected and untreated. The presence and subsequent progression of chronic kidney diseases are primarily measured by the extent of interstitial fibrosis and glomerulosclerosis in renal biopsies. The traditional method of interstitial fibrosis quantification uses visual evaluation, which is confounded by high inter- and intra-observer variability. This variability in diagnosis is an undesirable and potentially hazardous trait of human judgement. This thesis aims to minimise the human variability in the assessment of renal interstitial fibrosis by the introduction of a computerised quantification system.

To attain a higher accuracy in the deduction of interstitial fibrosis amount in renal biopsy tissue images, the identification and segmentation of all renal tissue structures were carried out in the proposed system. Computer vision rules for each tissue structure were explicitly set out based on interpretations of pathologists' judgement. A combination of colour spaces was also used to identify pixels of red and blue colour – the dominant colours in Masson's trichrome-stained biopsy tissues. Compared against conventional pure colour intensity thresholding and clustering methods, the proposed algorithm successfully halved the average error in interstitial fibrosis region segmentation. Of particular interest is the glomerulus segmentation, which remains an open research question due to the structure's large variation in appearance in the biopsy images. A novel textural feature descriptor called the channel correlation feature has been proposed

for use in the segmentation of healthy glomerulus structures. This feature is derived from the difference measure and correlation between multiple descriptions of the same image. It characterises the image texture and generates a feature that is invariant to scale, translation and rotation. When tested on a compiled glomerulus image dataset, it has achieved a 100% classification accuracy, while an average accuracy of 73.1% was recorded for the glomerulus extraction algorithm when tested on a 20-image renal biopsy dataset with manually annotated glomerulus regions. Due to the significant amount of possible variation in the appearance of kidney tissue structures in biopsy samples, there is a lack of an established ground truth image set in the literature on interstitial fibrosis quantification. This research project has therefore undertook the task of compiling a set of 40 images of biopsy images with annotated interstitial fibrosis regions.

In order to prove the feasibility of the automated quantification system as a diagnostic tool, a group of five experienced pathologists are involved in the experiment rounds. By incorporating the proposed automated quantification system for interstitial fibrosis into the diagnosis process of practising pathologists, the pathologists' intra- and inter-observer variability have been shown to decrease. With these objective quantification results as a reference, pathologists have shown an improved consistency and accuracy in their diagnoses as reflected by an increment of one agreement category (out of six) between the pathologists. This has thus fulfilled the primary research aim of producing an automated quantification framework capable of reducing the human variability in visual evaluations of renal biopsy images.

Publications

1. **W. K. Tey**, Y. C. Kuang, M. P.-L. Ooi, and J. J. Khoo, "Channel correlation image descriptor in cell segmentation," *submitted for publication in Pattern Recognition.*

2. **W. K. Tey**, Y. C. Kuang, M. P.-L. Ooi, and J. J. Khoo, "Automated quantification of renal interstitial fibrosis for computer-aided diagnosis: A comprehensive tissue structure segmentation method," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 109-120, 2018.

3. J. J. Khoo, **W. K. Tey**, V. Tan, S. W. Peter, Y. C. Kuang, and M. P. L. Ooi, "Visual Quantification of Renal interstitial fibrosis: Inter and Intra-observer variations," *Pathology*, vol. 49 (2017), p. S130.

4. **W. K. Tey**, Y. C. Kuang, J. J. Khoo, M. P.-L. Ooi, and S. Demidenko, " Automatic renal interstitial fibrosis quantification system," in *2017 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2017, pp. 1-6.

5. **W. K. Tey**, Y. C. Kuang, J. J. Khoo, M. P.-L. Ooi, and S. Demidenko, "Automating measurement of renal interstitial fibrosis: Effect of colour spaces on quantification," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2016, pp. 1-6.

Table of contents

List of figures

List of tables

Acronyms

ANN	Artificial Neural Network
ANOVA	Analysis of Variance
CAD	Computer-aided Diagnosis
CNN	Convolutional Neural Network
GLCM	Grey Level Co-occurrence Matrix
GLRL	Grey Level Run Length Matrix
GUI	Graphical User Interface
HOG	Histogram of Oriented Gradients
H&E	Hematoxylin and Eosin
ICC	Intraclass Correlation Coefficient
IF	Interstitial Fibrosis
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
MT	Masson's Trichrome
rLADTRee	Regularised Logistic Alternating Decision Tree
SFTA	Segmentation-based Fractal Texture Analysis

SIFT	Scale-Invariant Feature Transform
Sparse ADTree	Sparse Alternating Decision Tree
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

Chapter 1

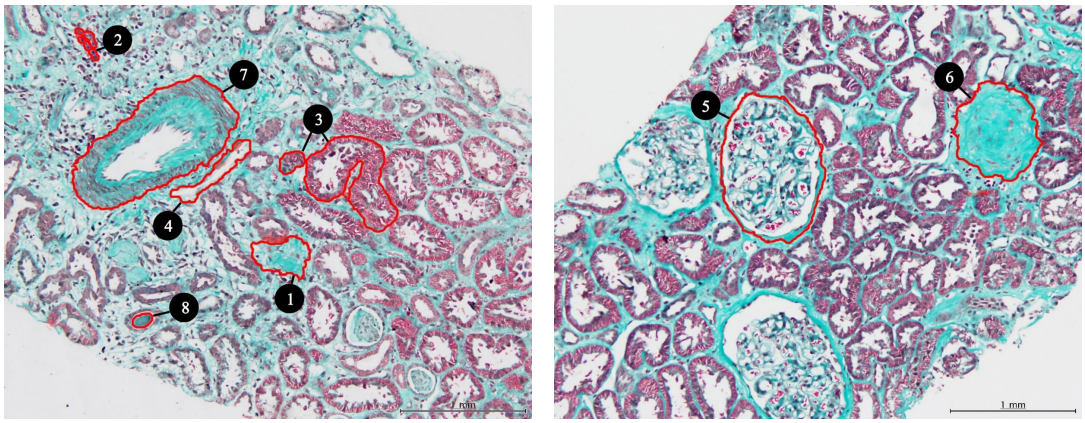
Introduction

1.1 Kidney tissue structures – An introduction

This work describes a computerised methodology to extract and quantify the interstitial fibrosis (IF) structure in kidney biopsy images, the first natural step is to identify the appearance of this structure in the image. However, the identification cannot be limited only to the IF structure alone. This is because the task of renal biopsy evaluation also requires the recognition of the various other structures typically seen in a kidney biopsy tissue as well, as their presence and condition could also serve as an indication to the kidney health. Therefore, a description of all the commonly seen tissue structures in a kidney biopsy tissue was compiled with the help of a trained pathologist, as presented in Table 1.1 and its corresponding example image in Figure 1.1. This comprehensive list of structures and their corresponding description of appearances provides a basis for subsequent structure segmentation and extraction algorithm choices.

Table 1.1 A visual description of all the common tissue structures in a Masson's trichrome-stained kidney biopsy tissue and their inclusion in the renal IF quantification, with a ✓ indicating the structure being counted towards the fibrosis region while a ✗ indicates the inverse.

Num.	Tissue Structure	Appearance in Masson's trichrome-stained biopsy image	Fibrosis Quantification
1.	Interstitial fibrosis	Green-blue region	✓
2.	Inflammatory cells sitting atop a plane of fibrosis	Small red dots surrounded by green-blue region with or without white spaces	✓
3.	Tubular structures	Thick red muscular walls encircling a blank white area	✗
4.	Lumens and cellular fluid spaces	Large white spaces	✗
5.	Healthy glomeruli	Circular structure with a highly textured region, surrounded by a circular gap region, including the fibrous region attached to it	✗
6.	Sclerosed glomeruli	Circular uniformly bright blue structures	✗
7.	Arteries	White space surrounded by strips of red muscular walls	✗
8.	Urinary casts	Blue or red structure regions enclosed by white boundary within red tubule walls	✗



(a) (b)

Figure 1.1 Sample biopsy images showing example regions of corresponding structures in Table 1.1 (a) Structures 1—4 and 7—8 (b) Structures 5 and 6.

1.2 Kidney disease and renal interstitial fibrosis

The kidney serves an important role in the day-to-day function of the human body. It is the filtering device in the human circulatory system and is responsible for removing waste products from metabolism and excess water from the blood. Kidney failure results in the accumulation of toxic waste in the body. If left untreated, this accumulation of toxins in the human body could be fatal. According to the 2010 Global Burden of Diseases study, the number of deaths from chronic kidney disease has seen a sharp increase between the years 1990 and 2010 which was only second to that for HIV and AIDS [1]. Unfortunately, this trend is expected to increase with the growth of the global elderly population, due to their increased susceptibility to diabetes, hypertension, and cardiovascular events [2].

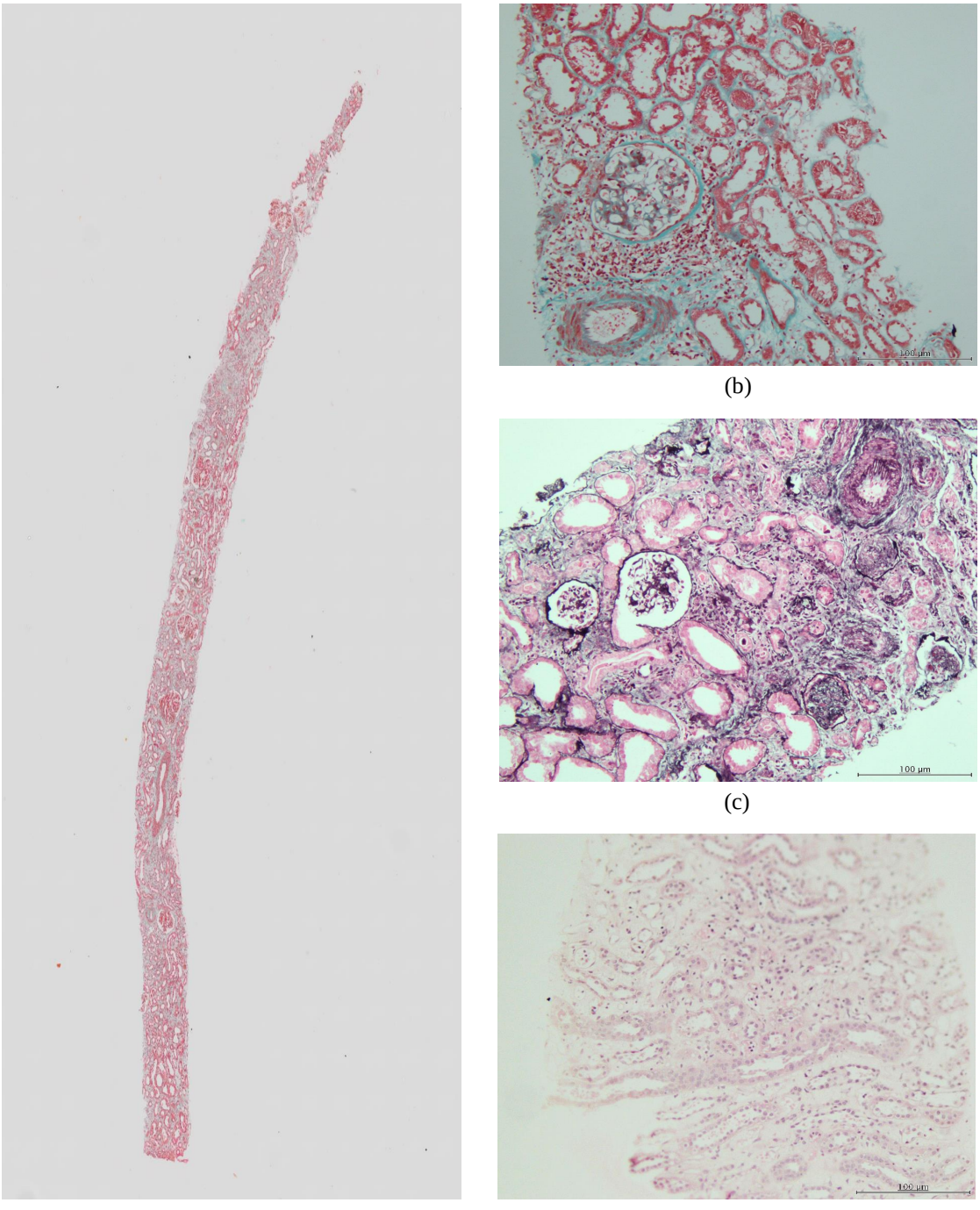
With early diagnosis and treatment, chronic kidney diseases can be controlled and its progression towards end-stage renal failure can be delayed. The first symptoms of a possible kidney disease are typically discovered through blood and urine tests. Early diagnosis to ascertain the presence and activity of the kidney disease requires renal biopsy samples to be taken and studied from the patient, and this practice is the

gold standard [3]. The biopsy samples taken are observed for the presence of specific pathological structures, whereby one of the main identifying structures of chronic kidney disease is the IF. An example of the renal biopsy tissue sample is shown in Figure 1.2 (a).

IF is the excessive accumulation of collagen in the extracellular region – the kidney tissue region that surrounds the large tissue structures (such as renal tubules, glomeruli and arteries). This is normally present in the tissue following the healing of a wound. In other words, IF is the scarring tissue which displaces healthy specialised tissues in the organ, thus causing a loss in functionality of the affected organ [4]. The exact description of IF and a sample image of it is provided in Figure 1.3. The amount of IF in an organ is usually an indication of its health, and this information is of specific interest in the kidney, where the extent of the IF progression has been shown to be the best indicator of end-stage kidney failure [5]. Therefore, the quantification of IF in renal biopsy samples by trained pathologists is a common practice in the diagnosis and prognosis of chronic renal diseases.

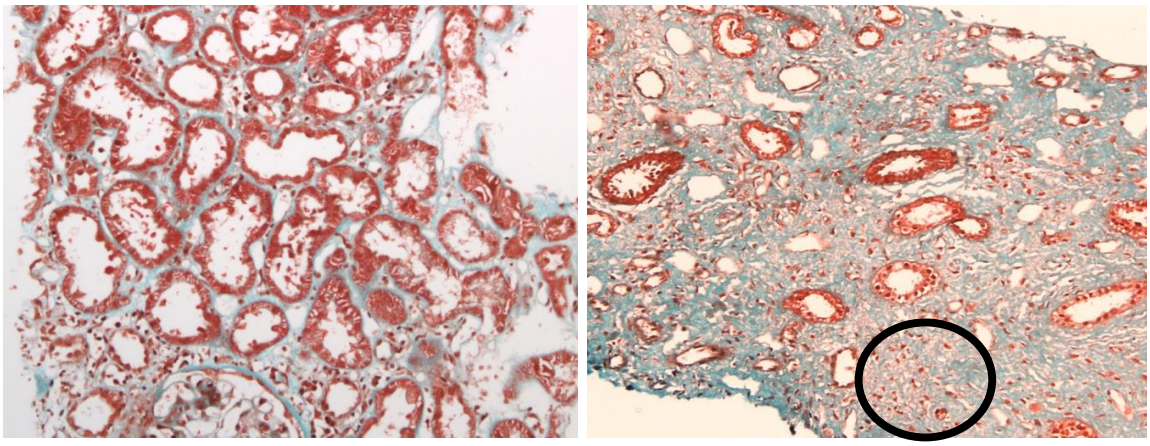
The conventional approach to renal IF quantification is achieved by staining the renal biopsy sample with chemical dyes to highlight the tissue structures of interest [6]. One of the most prevalent stains used for the inspection of renal biopsy samples is the Masson's trichrome (MT). It reacts with the collagen in the tissue to produce a blue-green colour, which is contrasted against a red background of muscular tubular structures [7]. Some common chemical stains used in histopathology besides the MT are the Hematoxylin and Eosin (H&E), Sirius Red and Periodic Acid-Schiff . Several examples of biopsy tissues stained by these different types of stains are provided in Figure 1.2, while Figure 1.3 shows samples of MT-stained biopsies of a (a) healthy and (b) diseased kidney. Upon the staining of the biopsy sample, a trained pathologist evaluates it under an optical microscope for structures or markers indicative of kidney disease.

The assessment of IF in biopsy samples by visual evaluation of the pathologists is considered the standard practice in the diagnosis process [6]. The estimated amount



(a) (b) (c) (d)

Figure 1.2 Samples of kidney biopsy images observed under an optical microscope, where (a) shows the entire Masson's trichrome-stained renal biopsy sample, while other smaller sections of the biopsy sample are stained by commonly used stains, such as the (b) Masson's trichrome, (c) Periodic acid-silver methenamine, and (d) Hematoxylin and eosin.



(a) (b)

Figure 1.3 Histopathological slides depicting (a) healthy kidney and (b) unhealthy kidney whereby an example of the renal fibrosis is shown in the black circle.

of IF is expressed as a percentage of the biopsy tissue area and is routinely compared against the Banff classification criteria for renal allograft diagnoses [8]. The Banff classification categories allow the pathologists to make recommendations and perform diagnoses in a semiquantitative manner. This type of generalised grading system for IF classification is also used in the diagnosis of IgA nephropathy [9] and lupus nephritis [10], which are diseases that cause inflammation in the kidney.

According to the Banff classification criteria, the severity of IF is graded in three categories with respect to the total cortical area. The biopsy tissues acquired for the analysis of IF are of the renal cortex – the outer section of the kidney that performs blood filtration. Hence, the total biopsy area is equivalent to the total cortical area in this study. The Banff categories are:

I. mild (IF $<$ 25% of cortical area),

II. moderate (IF 26 — 50% of cortical area), and

III. severe (IF $>$ 50% of cortical area).

Sample biopsy tissues classified under each of the three Banff categories are presented in Figure 1.4.

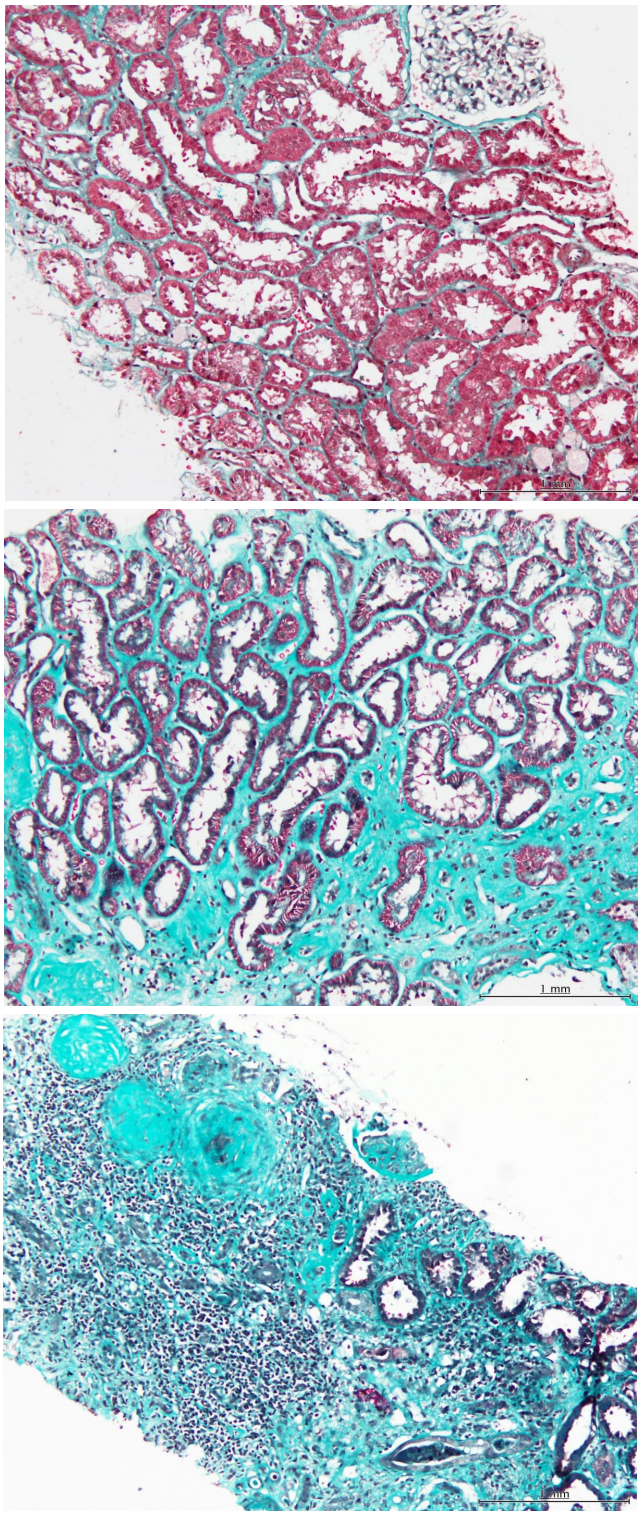


Figure 1.4 The visualisation of the Banff categories as demonstrated in sample renal biopsy images.

1.3 Problem definition

In this project, the core focus is on the assessment of renal IF in MT-stained kidney biopsy samples. The conventional way of producing a diagnosis for renal interstitial fibrosis involves the pathologists estimating the fibrosis percentage in a biopsy image based purely on their experience and educated judgement. As described previously, the Banff criteria is a guideline for grading the severity of kidney diseases by visual evaluation of biopsy samples. Since the Banff criteria is a loose guideline based on human judgement, the quantification of renal IF by visual evaluation is susceptible to intra- and inter-observer variability. Addressing the poor repeatable or reproducibility of the visual evaluation of renal biopsy images [11–13] is an important factor that improves the quality of medical care.

Advances in computerised tools in the biomedical area has spurred the development of automated image analysis systems to aid pathologists in accurately identifying human tissue structures [14]. This solution can be applied to the problem of IF quantification. Due to the automated and computerised nature of quantification, inter/intra-variability of pathologists in the assessment of the extent of fibrosis can thus be significantly reduced to give a more consistent result. This will in turn produce repeatable and reproducible analyses of chronic renal diseases, and ultimately improve the diagnosis and prognosis of kidney diseases. Therefore, the feasibility of building a computerised workflow for IF quantification is examined in this thesis.

1.4 Research motivations and aims

The task at hand is an interdisciplinary one, bridging the gap between a real medical need and a feasible engineering solution. Aided by the explosion of digital medical information made available in recent years, the idea of computer-aided diagnosis systems

is growing into a commonplace reality in many medical and healthcare settings. From an engineering and computer vision standpoint, a system that is capable of robustly quantifying IF while taking into account all commonly encountered kidney tissue structures has yet to exist.

At the present, one system that comes the closest is that proposed by Meas-Yedid et al. [15]. The working framework of the quantification system is based on the premise of the detection of blue-green coloured pixels in the biopsy image, and subsequently eliminating non-fibrosis blue-green structures from the detected regions such as the renal capsule, tubular basement membranes and sclerotic glomeruli. This has been the classic approach towards the segmentation of IF in MT-stained biopsy images, as the natural contrast presented by the stain eases the identification of collagenic fibrosis structures. Nevertheless, this detection method's dependence on the colour contrast alone creates a problem in old biopsy samples with faded staining and colour variations due to material, lighting and imaging conditions. A review of quantification systems utilised on biopsy images, not only of the kidney tissue but other organs as well, is included in Chapters 2.7 and 2.8.

Hence, instead of grounding the focus on blue-green tissue sections, this thesis work explores the tissue structure segmentation from the perspective of red tissue structure detection. The rationale behind this decision is that the intensity of the red stain is significantly better preserved over time as compared to the blue-green stain. Moreover, pathologist feedback has suggested that certain small red structures, traditionally ignored by computerised system, should be included in the area classified as IF for a more comprehensive quantification, as these structures are actually sitting atop the plane of IF. Therefore, it is proposed to work around this issue by focusing on the detection of red-stained pixels signifying keratin and muscle structures for removal from the total tissue region, combined with a textural feature validation on the selected regions. Additionally, glomerulus segmentation is still a topic under popular research

due to its variability in appearance in a biopsy sample. Therefore, this thesis also delves into the search for a novel solution to this problem. Finally, the proposed framework was tested on a set of renal biopsy images with ground truth prepared in collaboration with an expert pathologist.

The aim of this research is to create a methodology to quantify the renal IF in kidney biopsy images and subsequently develop and validate a clinical renal fibrosis quantification tool that adheres to the general consensus of the pathologists.

To achieve this aim, several objectives must be met:

1. Investigate the inter- and intra-variability in a group of trained pathologists on IF quantification by visual evaluation of kidney biopsy images.

2. Identify all the common kidney tissue structures (including IF) through automated segmentation.

3. Develop segmentation algorithms to extract the regions for each structure of interest, combining methods of feature detection, extraction, and classification in machine vision.

4. Validate the performance of the quantification system for renal IF by statistical analysis and the generated ground truth image dataset.

5. Assess the applicability of the finalised automated quantification system as a diagnostic aide to pathologists in their visual evaluation process.

1.5 Overview of contributions

Under the main objective of designing and formulating a framework for the systematic quantification and segmentation of IF in renal biopsy images, this research has

accomplished several subsidiary goals. The key contributions of this research project are listed as follows.

I. The introduction of an automated workflow for renal IF quantification through the explicit definition of computer vision and image processing rules for the segmentation of each renal tissue structure.

II. The development of a glomerulus structure detection and segmentation method by using a novel image textural classification method.

III. The analysis on the variability in the pathologists' standard practice of visual evaluation of renal biopsy samples and the impact brought forth by the use of the proposed automated quantification tool.

IV. The compilation of a 40-image biopsy dataset with manually annotated ground truth of IF regions in a range of diseased states of the kidney tissue.

The primary output from this research is the development of a complete algorithm for the detection of all common kidney tissue structures found in renal biopsies images, towards the quantification of IF. The appearance of each structure in a biopsy was defined in computer vision terms built upon the implicit observation rules held by a practising pathologist in the visual evaluation of renal biopsies. The segmentation technique for each tissue structure employs a combination of colour and image texture information, morphological operations, and geometrical shape fitting. The main distinction of the proposed system from the others available in the literature is in its approach towards the inference of IF areas in the biopsy. Specifically, the segmentation method is built around the recognition of red renal structures rather than the blue-green ones. The performance-validating experiments were also executed on a manually prepared ground truth image dataset with annotated IF regions.

One renal tissue structure that is given particular attention in this thesis is the glomerulus. This structure is ubiquitous in the renal cortex region that it is a defining feature in renal biopsy tissues. Reliable identification of the glomeruli is critical to the accuracy of the IF quantification. For the purpose of glomerulus segmentation, a novel image classification feature based on the similarity and correlation measure between different image representations was proposed. Examples of the image representations tested are greyscale and colour intensities, wavelet decompositions, and gradient magnitude images. This classification feature was tested on a set of manually cropped glomerulus images and has achieved excellent results.

Given that the pathologists are the intended beneficiaries of the proposed quantification system, this research project necessarily performed experiments involving a group of five pathologists to examine their diagnosis procedure. The extent of intra- and inter-pathologist variability was determined empirically through a series of quantification experiments requiring participants to evaluate sections of renal biopsy images. Ultimately, the validation of the developed automated IF quantification system as a diagnostic aide to the pathologists was performed. The proposed system has shown a marked positive impact on the agreement between pathologists in their diagnosis of the amount of IF in renal biopsies.

To allow for an accurate measure of the performance of the proposed quantification system, the preparation of a ground truth dataset is crucial process. There is a lack of an established ground truth image set in the literature on IF quantification due to the significant amount of possible variation in the appearance of kidney tissue structures in biopsy samples. This is especially evident in severely atrophied or diseased tissue samples. For example, in a similar research on liver biopsy tissue images [47], the evaluation dataset of 79 images had to be manually selected and annotated by experts. However, this research undertook the task of compiling a 40-image ground truth dataset of renal biopsy images with manually annotated regions of IF and other relevant tissue

structures with the aid of an experienced pathologist. This was an essential step to carry out the validation and analysis on the accuracy of the segmentation algorithms. This 40-image dataset is a subset of the biopsy images collected from 70 patients that encompasses the range of healthy to diseased renal tissues. The dataset is the most comprehensive to support future development in the field.

1.6 Thesis organisation

Chapter 1 Introduction

This chapter gives an introduction to the dissertation and definition of the area of research. It presents a description of the appearance of each renal tissue structure commonly found in the biopsy. An overview of the research problem along with the achievables and expected outcomes of the research project are also succinctly listed.

Chapter 2 Interstitial Fibrosis Quantification: A Review of Approaches and Techniques in Medical Image Analysis

The chapter starts out with a comprehensive review on current computer-aided diagnosis systems in various organs of the human body to examine the different algorithms and process flows involved in the designs. Next, a thorough examination on the image segmentation and classification techniques implemented in the medical setting is provided as background for the research work in this thesis. Then, the state of the research on glomerulus image segmentation from biopsy images is established with examples from the literature. Finally, an overview of existing IF quantification systems (involving both renal and other organs' biopsy samples) are presented to identify the research gap in the relevant area.

Chapter 3 Knowledge-based Renal Interstitial Fibrosis Quantification

This chapter details the experimentation of the use of different colour spaces in renal tubular structure segmentation and compares between single and combination colour space methods. The results form the basis for the identification and segmentation rules for each tissue structure, which utilises various image processing techniques. As a benchmark procedure, the proposed quantification method is evaluated against the established method of IF quantification by colour thresholding and clustering.

Chapter 4 Segmentation of Glomeruli and Complete Quantification System Analysis

The segmentation of the healthy and sclerosed glomerulus is explored in a separate chapter to allow for a more in-depth investigation. Various methods for healthy glomerulus segmentation are experimented and contrasted against each other in the search for the optimal technique for glomerulus segmentation. A new method which exploits the relationship between different image representations is proposed as a glomerulus image classifier, among other established techniques involving morphology matching and image texture classification. Finally, the performance of the complete quantification system encompassing the segmentation of all renal tissue structures is inspected and analysed.

Chapter 5 Clinical Application

In this chapter, the outcome from experiments conducted with practising pathologists are detailed. The extent of variability in the standard procedure of IF quantification by visual evaluation is measured to bring context to the performance of the developed automated quantification system. Lastly, the automated quantification system is put to test as an aide in the pathologists' evaluation of the biopsy samples. A prototype of the graphical user interface

developed for the IF quantification system is also described in terms of its functionality and aspects for improvement.

Chapter 6 Conclusion

A summary of the main findings and contributions of this dissertation is covered in this chapter. A brief discussion on the research topic and probable extensions to the work in the future wraps up this thesis.

Chapter 2

Interstitial Fibrosis Quantification: A Review of Approaches and Techniques in Medical Image Analysis

2.1 Introduction

IF quantification is one of the most indicative test procedure employed to examine the progression of chronic kidney disease [5], and is traditionally carried out by visual evaluation of skilled pathologists. As stated in Chapter 1, the gold standard for the diagnosis of IF in kidney samples is the Banff classification system, which sets up criteria for diagnosing the severity of kidney disease based on the corresponding amount of IF found in the kidney sample. The creation of this standard is exactly an attempt to overcome the inherent human limitation to provide a precise measurement in a complex biopsy image. In recent years, the introduction of computer-aided diagnosis (CAD) systems has provided a distinct solution to the problem through the use of computerised and automated tools. A well-designed CAD system can achieve better performance

consistency, higher speed, greater accuracy and lower operational cost, which are all very important to the improvement of medical care quality.

Therefore, this chapter endeavours to paint a broad picture of the process by which a typical computer-aided visual diagnosis system is set up, with the flow illustrated in Figure 2.1. The chapter starts out with a review on the current CAD systems targeted for detecting diseases in various organs in Chapter 2.2. This is followed by a summary of the commonly utilised image features for structure segmentation in Chapter 2.3, the popular classification methods in Chapter 2.4, and other general image processing methods in Chapter 2.5. As the glomerulus segmentation problem is still an active research topic, Chapter 2.6 describes some of the current solutions to glomerulus detection and identification in biopsy images. A brief compilation of the up-to-date computerised quantification systems developed for detecting fibrotic structures is provided in Chapter 2.7. Finally, Chapter 2.8 focuses on the automated quantification systems for renal IF that exist in the literature.

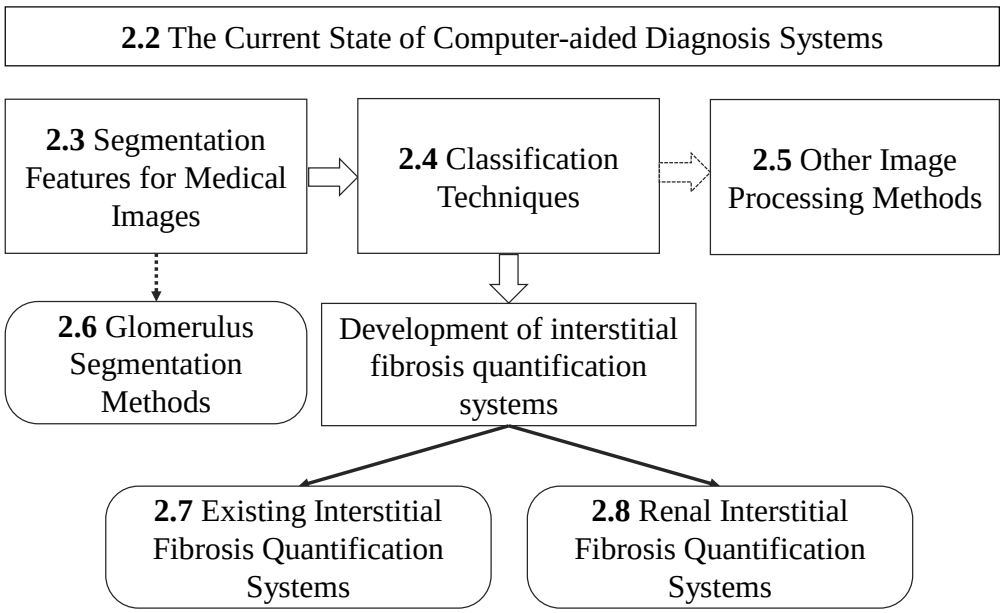


Figure 2.1 The flow of the subsections in Chapter 2, which mimics the typical process flow for the development of a computer-aided diagnosis system, with specific focus on renal biopsies.

2.2 The current state of computer-aided diagnosis systems

The advent of computers with high processing capabilities has facilitated the design and use of CAD tools in the biomedical setting. CAD systems are computer programs designed to serve as a "second opinion" to medical experts in their diagnoses by automatically detecting potential regions or structures of interest in medical images. This helps the experts to perform their analysis on the data quicker and have more confidence on their diagnosis. In fact, the aid of an automated tool has been proven to increase the cancer detection rate in mammography by 10%, on average [16]. All subsequent discussions will only involve image-based CAD systems for its direct relevance to the project. This is necessary to limit the scope of review because CAD systems in general span a large breadth across many disciplines and many different technologies.

Combining CAD with imaging systems is increasingly commonplace in the healthcare industry as technology advances and cost decreases [17]. The precision and consistency of the computer system provide an advantage over human judgment in medical image analysis, especially in the detection and classification of unusual or malignant tissue structures in digital images [18]. Having a high accuracy and consistency is vital in the medical diagnosis process. This trait is especially desirable in routine high-volume medical screening and diagnostic procedures as they are labour intensive activities. Aside from the skill and experience of the human observer, the accuracy of the diagnosis may also be affected by human conditions such as fatigue, lack of focus and the emotional state of the observer [19].

As a rule, the general framework for developing a CAD system consists of a pre-processing stage, a segmentation stage, a feature extraction stage, and a classification stage, as illustrated in Figure 2.2. Feature generation and classification stages cover a large group of techniques that will be elaborated separately in Chapter 2.3 and 2.4.

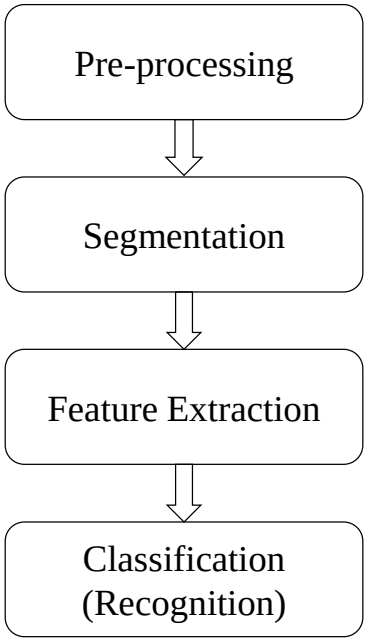


Figure 2.2 The general framework ubiquitously adopted for designing a computer-aided diagnosis system with a dataset of digital biomedical images.

The existing CAD systems are typically task-specific and can be grouped based on the organ they were developed for. There are large combinations of feature generation and classifier for each organ in the literature. The breakdown according to Figure 2.2 allows for easier overview and understanding of the techniques used.

The organs that are under study for CAD include the breast, chest, skin, colon, brain, liver, kidney, vascular and skeletal systems. Based on a recent bibliometric review performed by Takahashi et al. [20], breast cancer is the subject of the highest amount of paper published on CAD systems, followed closely by pulmonary disease. This comes as no surprise as the earliest CAD systems were developed for mammograph images in diagnosing breast cancer [21] and chest radiographs in detecting heart abnormalities [22, 23]. The following review compiles the CAD systems developed and found in the literature, based on the organ they were designed for.

The method for evaluating the effectiveness of a CAD system is split into three phases, as suggested by the paper by Takahashi et al. [20]. The process starts with the building

of the image database, before the evaluation of the computer system and finally, the receivers, who are the pathologists in this research. Indeed, these core components of a CAD system have been found to be present in most CAD systems across all organs. Therefore, this research emulates the same course of evaluation for the validation of the proposed quantification system, as detailed in subsequent chapters.

As an overview of the CAD systems for the organs investigated in this chapter, the CAD systems for the breast, chest and colon are among the more well-explored ones. They share a common characteristic in that the pixel intensity feature is heavily involved in most of their feature extraction algorithms. It is also usually combined with textural and geometrical features for structure segmentation. In contrast, CAD systems for the skin focus mainly on the detection of edge features. The reason behind it is that the skin CAD systems mostly target the detection of skin lesion boundaries, as well as hair artefacts, which appear as hard lines and edges. However, for the brain, the classification of an ensemble of features takes up the bulk of the research on the CAD systems for the organ.

2.2.1 Breast

CAD systems for breast cancer are perhaps the best explored among all CAD systems for human organs. The various imaging modalities that exist for diagnosing breast cancer is an indirect testament to the long-standing history in the usage of technology for its diagnosis. Some examples of the imaging techniques applied for diagnosing breast cancer are digital mammography, ultrasound, and magnetic resonance imaging. These modalities produce distinct images that highlight various abnormalities that can help a radiologist diagnose for breast cancer. Nonetheless, the steps for producing a diagnosis using these images still largely follow that which was described in Figure 2.2.

The detection of breast cancer is usually carried out in terms of the detection of abnormalities or diagnostic features in the biomedical images. The types of features looked for are microcalcifications, masses, bilateral asymmetry and architectural distortion. The detection of microcalcifications and masses in breast images is a well-established method for breast cancer diagnosis, while more subtle cues that might indicate an abnormality in the breast remain a more niched research area.

Microcalcifications are deposits of calcium that show up as brighter regions than the surrounding breast tissue in a mammogram. Malignant microcalcifications are usually small, numerous and appear in clusters [24]. CAD systems based on microcalcification detection generally exploit its characteristic property of being brighter than its surrounding, or its statistical difference with the background tissues. Examples of methods employed are contrast-enhancement for noise removal as a pre-processing step [25, 26], higher order statistics modelling on image channel pixel intensity [27, 28] and multiscale wavelet decomposition methods [29] for feature extraction, and various machine learning classification techniques such as support vector machine (SVM) and artificial neural networks (ANN) [30].

On the other hand, the localisation of masses or tumours in mammographs have a more straightforward approach of detection and classification. Features are first derived from the images in search of suspicious regions before being classified by a statistical model for validation. Since the shape and appearance of the mass decides its malignancy, geometrical [31] and image texture analysis [32] are among the popular approaches. The literature is especially directed at the implementation of textural features for tumour classification, such as multiresolution wavelet features [33], grey-level dependence matrices [34] and grey-level co-occurrence matrices (GLCM) [35].

Architectural distortion in the breast tissue with no visible mass (including spiculations radiating from a point, and focal retraction and distortion at the edge of the parenchyma) is one of the often missed cue for breast cancer diagnosis [36]. The distortion in the

breast tissue is proposed to be detected through textural description of the tissue region such as Gabor filtering [37], texture orientation and flow-field [38], and structure morphology [39]. In particular, the flow-field depicts the orientation of the fibro-glandular tissues in the mammogram, which characterises the anisotropy of the image texture. Another observed clue towards breast cancer diagnosis is the bilateral asymmetry in mammograms. A difference in the appearance between the left and right breasts may indicate the presence of breast cancer at an early stage [40]. The asymmetry may be in the form of volume of the breast tissue, density of the tissue, and the prominence of ducts. From the few research papers addressing this task, the asymmetry analysis is achieved by incorporating measures of shape, texture, density, brightness, directionality [41–43].

2.2.2 Chest

The commonly utilised image acquisition technique in chest scans are chest X-ray [44], computed tomography (3D cross-sectional images from 2D X-ray images) [45], positron emission tomography [46], and ultrasound [11]. With computerised tomography being the leading choice for assessing lung diseases due to its high level of contrast, resolution and detail, the literature for chest scans are focused on using computed tomography scans as the image database for CAD systems [47]. With this imaging modality commonly included as part of the diagnosis process in hospitals, CAD systems in this area mostly revolve around the detection of pulmonary nodules and the diagnosis of interstitial lung disease. Other less investigated applications of chest scan CAD systems are in the detection of cardiomegaly [48], pneumothorax [49], and tuberculosis [50].

Pulmonary nodules are small masses of tissues in the lungs and an important diagnostic feature for lung cancer. These nodules appear as brighter round shadows in the X-

ray images of the lung. The challenge encountered in this image database lies in distinguishing between the nodules and the ribs and vessels, which are often overlapping in the radiographs [51]. Proposed CAD systems for the detection of pulmonary nodules make use of the image pixel intensity values [52] and cell morphology [53, 54] to segment these structures, while image textural features are often employed to eliminate false positives in the detection. Some of the textural features used in this task are co-occurrence matrix-based features [55], Laplacian of Gaussian [56], and Gabor kernels [57].

Conversely, CAD systems designed for diagnosing interstitial lung diseases puts more emphasis on textural features than morphological features. This is because the manifestations of the disease are more diverse due to the more complex background of the computed tomography image [58]. Several CAD systems have made use of the multi-scale texture analysis in building a model for the classification of interstitial lung disease through the use of multi-scale filter banks and wavelet analysis [59]. Furthermore, the adoption of a combination of multi-scale features and the more basic grey level features have resulted in an improved sensitivity of the radiologists in their diagnosis, as recorded by the receiver operating characteristics analysis [60].

2.2.3 Colon

With colorectal cancer being among the top three most commonly diagnosed cancers worldwide [61], CAD systems surrounding the disease that exist are numerous. Computer tomography colonography is a non-invasive screening technique equivalent to a virtual colonoscopy for which most CAD systems in this area of research is built. Computer tomography colonography allows the search for precursors or indicators of colorectal cancer in the colon, such as polyps and masses, in a safer manner through the 3D model of the patient's colon [62].

The detection of polyps in radiological images begins with the detection of the colon itself, as the polyps are growth that are attached to the walls of the organ [63]. This step is achieved by exploiting the computed tomography intensity values and applying a region growing method for the extraction of the colon from other organs with soft-tissue density [64]. Upon the segmentation of the colon images, the potential polyps are detected based on the characteristic shape and geometry of the polyps, which present themselves as a bulbous and cap-like structure. The curvature of the colonic wall is quantified in terms of a shape index for identifying the polyps growing on the walls [65]. Other methods of detection include the extraction of grey-level-based features and texture features to be classified by a machine learning model such as an SVM [66, 67], linear discriminant analysis (LDA) [68], and massive-training artificial neural network [69, 70]. The massive-training artificial neural network is a pixel-based ANN trained on overlapping image sub-regions and a "teaching" image of the likelihood of each pixel region being a lesion.

The reduction of false positives plays a big part in the workflow of the colonography CAD system, as with any other biomedical CAD system. The main sources of false positive come from thickened haustral folds and retained solid stools in the colon. They can be eliminated by a cross check on the location and appearance of these structures [71]. On the other hand, colorectal masses are structures that are easily identifiable by radiologists but often misclassified by CAD systems as multiple polyps. The reason for this is the wide variation in shape characteristics of the masses [72]. Among the little research work investigating colorectal mass detection, a merging method based on fuzzy membership and a wall-thickening analysis based on computed tomography intensity values have been proposed by Näppi et al. [73] to distinguish between polyp and mass candidates.

2.2.4 Skin

The skin is the largest organ in the human body with the vital role of protecting the body from its harsh environment, especially solar ultraviolet rays. Excess ultraviolet rays can damage the deoxyribonucleic acid in the skin, causing mutation which results in an abnormal growth of melanocytes known as melanoma [74]. The cancer manifests itself as pigmented skin lesions, identified as discolouration on the skin that can vary in shape, size, degree of colouration, and symmetry [75]. The current scoring systems for skin lesions used by radiologists to determine if a lesion is malignant or benign still suffer from a low level of concordance between observers [76]. Examples of the scoring system are the ABCD-E rule [77], seven-point checklist [78], three-point checklist [79] and Menzies method [80].

CAD systems for diagnosing melanoma are largely based on dermoscopy and photography image datasets. There are four main stages in the CAD systems, namely image pre-processing, segmentation of lesion border, feature extraction, and feature classification. Skin lesion images are often improperly contrasted due to the variations in skin tones and can include hair artefacts that obstruct the proper segmentation of the lesion [81]. Therefore, images are usually pre-processed with contrast enhancement in terms of different colour space channels [82, 83]. Meanwhile, hair artefacts are eliminated using Radon transform (to detect the predominant orientation of the hair strands) and morphological operations (to extract hair contour information) [84, 85]. Subsequently, the lesion is extracted by various low-level and high-level segmentation techniques. Examples of the existing work use thresholding [86, 87], region merging [88, 89], edge-based [90, 91] and morphological segmentation methods [92, 93] to detect lesion boundaries. There has also been a number of recent works that fuses different segmentation techniques for lesion detection [82, 94, 95], as well as exploiting the use of deformable models as the lesions are generally of a circular geometry [96, 97].

Once the lesions are segmented, features of the lesion that are able to describe dermoscopic features such as pigment pattern and vascular pattern [98, 99] can then be extracted for classification. These features include shape [100, 101], colour [102, 103], texture [104, 105], and geometrical features [106]. They are able to quantify the lesion information that correspond to those looked for by radiologists in their visual evaluation. This stage is followed by classification of the dermoscopic lesion image based on these extracted features. The classifiers that have been employed for this task include the SVM [107, 108], ANN [109, 110], decision trees [111, 112], and ensemble classifiers [113, 114].

2.2.5 Brain

Brain tumours are the result of abnormal growth in the brain, which can be classified as benign or malignant based on the aggressiveness of the growth. These tumours can be present in a variety of sizes, shapes, and locations in the brain, while showing up in different intensities depending on the imaging conditions [115]. The most popular imaging techniques for diagnosing brain tumour are the magnetic resonance imaging and computed tomography due to their widespread availability and high resolution [116, 117]. Similar to other CAD systems, the brain CAD systems follow the flow of steps described in Figure 2.2.

Firstly, noise in the image is generally dealt with in the pre-processing stage using a median filter [118], mean filter or Gaussian filter to smoothen the image [119]. This step removes the high frequency noise in the image while increasing the image contrast to aid in the segmentation. Next, the segmentation of the brain tumour region is carried out by a wide variety of methods that ranges from simple thresholding and clustering to machine learning-based methods built around image intensity [120–122]. This is followed by the extraction of features to classify whether a tumour region is

malignant or benign [123]. The most popular features extracted in the literature include textural features [117, 124], wavelet transform features [125, 126], and shape features [127]. Finally, the classification of the features is performed through the use of machine learning classifiers such as the SVM [122, 128], ANN [129–131], and genetic algorithm [132, 133]. The genetic algorithm is a method of optimisation that is built around the principle of natural selection, which has been applied in feature extraction and selection procedures as well as classification methods [134]. However, ensemble classifiers appear to produce the best results for this application [117, 135].

2.3 Segmentation features for medical images

Histopathology is the study of a biological specimen for effects of a disease under a microscope. In histopathological images, tissue structures of interest are identified by their salient features. The reaction between chemical dyes and specific tissue structures in the specimen highlights them with distinct colours to facilitate visual judgement by the pathologist. Having the images corresponding to these structures to be quantitatively described in terms of a feature space allows for a measurable comparison between different structures.

In the context of CAD systems, feature extraction is an integral step in the detection of structures of interest prior to the recognition of the structures for a diagnosis to be made. These extracted features can be largely described by the following broad categories:

i. Colour / Grey Level Intensity

ii. Thresholding

iii. Edge Detection

iv. Region processing

v. Textural

vi. Geometrical Features and Shape Descriptors

vii. Graph-based / Architectural

It is important to make an important disclaimer at this point that, the above list of categories is not cast in stone. They are certainly not mutually exclusive as there are significant overlaps of concepts and techniques. Nevertheless, the categories are useful in aiding the appreciation of the dominant characteristics of each feature category.

The importance of these features is cemented by the availability of commercially available software made for image processing that implement the extraction of these features. For example, ImageJ [136] is a popular image processing program developed in the Java language used to process medical images. Its functionality includes the manipulation of colour information, edge detection, and thresholding of pixel values. Another software that exploits image features for segmentation is the bioimage analysis software called Icy. It allows the extraction of textural and shape descriptors in biomedical images for detection of biological structures. In the paper by Maree et al. [137], Icy is the platform used for the implementation of glomerulus detection in renal biopsy images. Furthermore, EIKONA3D [138] is a commercial software developed for interactive image segmentation. It also applies thresholding, region-based methods and edge detection for carrying out image object segmentation tasks.

2.3.1 Colour / Grey level features

The utilisation of colour space transformations in the domain of medical image segmentation has been well explored in the literature [139]. Colour carries a lot of information

in the stained histopathology images as explained in Chapter 1.2. Hence, the extraction and analysis of colour information plays an important role in this work. Greyscale forms the simplest digital image representation in which only the single-channel intensity values are recorded. Colour images are represented by a combination of three or more greyscale channels. The meaning of each greyscale channel differs from one colour space to another.

Colour spaces are essentially different representations of the colour and brightness information present in an image, made up of greyscale images that each represents one of the components. Colour images are conventionally represented in the *RGB* colour space, with three channels representing the *Red*, *Green* and *Blue* colour intensities respectively [140]. Unfortunately, this colour space is inherently unable to separate the chrominance component from the luminance component of an image [141]. Therefore, colour space transformations are performed in digital image analysis applications to represent the colour information in a way that facilitates processing. Figure 2.3 (a—g) gives an example on the contrast of colour information provided by the *RGB* and *YCbCr* colour spaces. Different colour spaces highlight different aspects of colour information that suit specific applications. For example, Dong et al. [142] introduced a new method in analysing ovarian cancer immunohistochemical images by utilising the U and V channels of the YUV colour space. The colour space use is justified by the fact that the positive cells and negative cells in images are distributed across separate regions of the U and V channels. A similar investigation is presented by Mao et al. [143] in the *YIQ* colour space, in which the I (hue) and Q (saturation) components contain the colour information of the image. Sadr et al. [144] introduced a segmentation algorithm for nucleus of leukocytes (white blood cells) that combines computations on the *YCbCr* channels and an active contour algorithm on blood smear slide images. Whereas in a paper by Liu et al. [145], tumour immunohistochemical pathology images are examined in the *YCbCr* colour space to identify certain cell components that are critical in the diagnosis for further treatment.

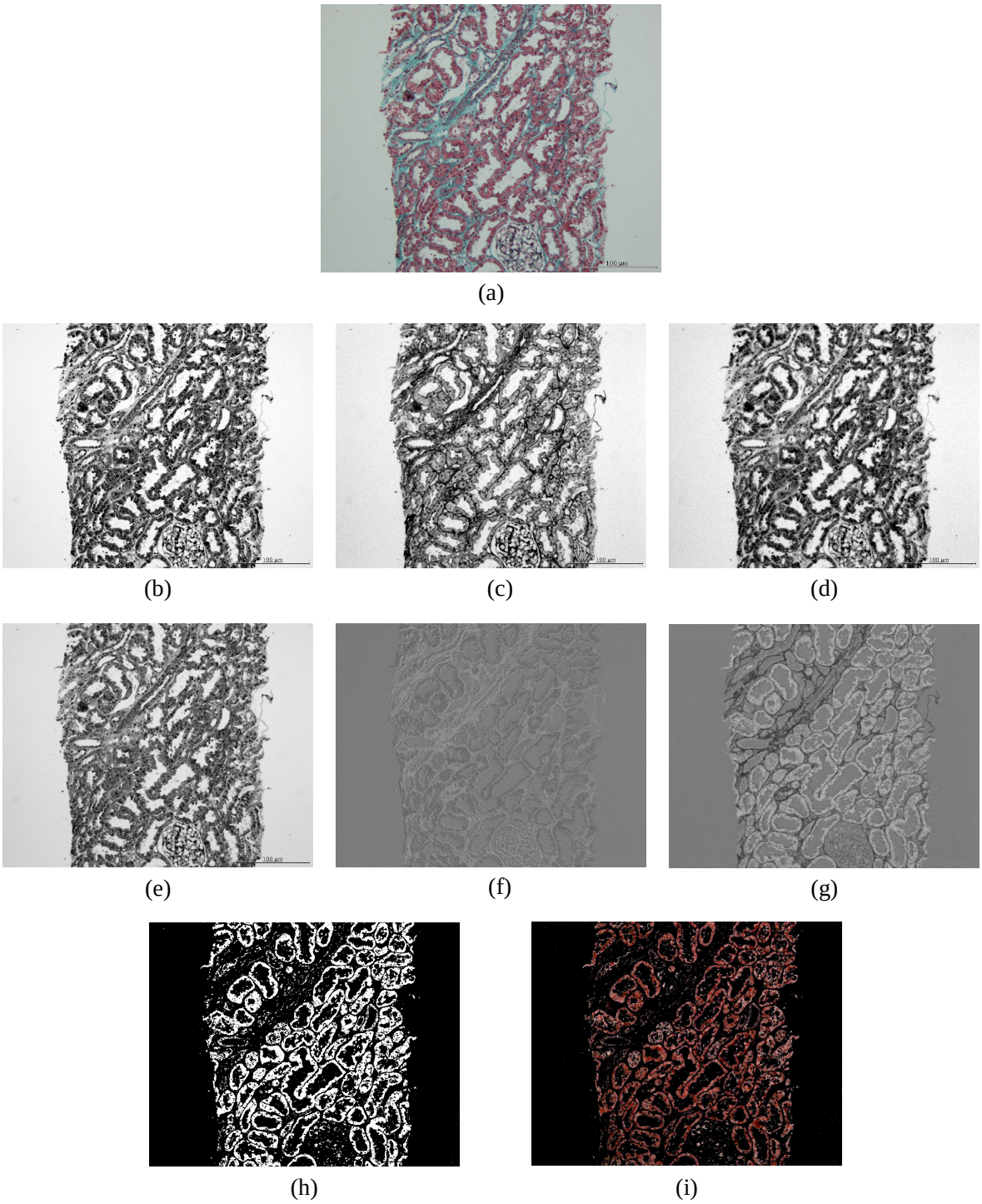


Figure 2.3 A visualisation of the information offered by different colour spaces based on a sample biopsy image as shown in (a). The corresponding intensity images of the *RGB* and *YCbCr* colour spaces are given in terms of the (b) *R*, (c) *G*, (d) *B*, (e) *Y'*, (f) *Cb*, and (g) *Cr* components. A simple global thresholding operation applied on the *Cr* component image produces an approximate binarised image of tubular wall structures in (h) and the complementary segmented structures in (i).

On the other hand, most of the research that makes use of the *HSI* colour space in histopathology are on the segmentation of white blood cells and their associated structures. This includes the extraction of the nucleus, cytoplasm and blast (abnormal white blood cell) [146–152] (similar to that by Sadr et al. [144], but no comparison between the two methods has been undertaken). All results allude to the fact that the contrast between the white blood cell and its background can be effectively enhanced by inspecting the hue (H) and saturation (S) components. Another work to be mentioned is done by Raof et al. [153], in which sputum slide images stained by Ziehl-Neelsen are analysed for tuberculosis bacilli. A comparison between *RGB* and *HSI* colour space thresholding methods is carried out, in which the *HSI* colour space outperforms the *RGB* colour space. A comparable *HSV* colour space was employed by Giotis et al. [154] to segment skin patches containing melanoma by implementing a k-means clustering on colour space components, while the *I1I2I3* colour coordinate system was used by Dhawan et al. [155] to enhance texture features for segmentation of skin pigmentation. Wang et al. [156] also presented a feature extraction method of tongue images in the *CIE XYZ* colour space based on its intuitionistic representation of visible colours.

The use of colour spaces for segmentation in digital medical images is not limited to single colour spaces or well-established ones. Dai et al. [157] have attempted to segment a range of hematocyte components by a combination of components from the *RGB*, *CMYK* and *CIE Lab* colour spaces. In another work, the *CIE Lab* colour space is exploited along with the *YCbCr* colour space to quantify bacilli and clusters in digital images of sputum smear samples [158]. Kong et al. [159] proposed a new *Most Discriminant Colour* space capable of maximising the discrimination power of local Fourier transform texture features generated. There are also research which include an ensemble of features produced from a large number of different colour spaces to be used in training a classifier [160, 161], whereas some attempt to implement a comparative study between the effectiveness of each colour space in the segmentation of their regions of interest [162, 163]. Due to the fact that combinatory colour spaces provide the

highest amount of corroborating information on the images, this research explores this method in the segmentation algorithms for red structures, specifically.

When the relationship between colour and the targeted object is stochastic in nature, the relationship between the pixels can be quantified statistically to be extracted as image classification features. Examples of low-level grey-level features based on histogram statistics that have been utilised on medical images are the moments of the distribution (mean, variance, skewness and kurtosis), entropy, correlation, and contrast [164].

2.3.2 Thresholding

Thresholding is a greyscale channel processing technique where intensity values are converted into binary values. The new binary-values image often contains valuable clues that can be used by the classification algorithms. Threshold-based methods are commonly used in the segmentation of light foreground objects in a darker background image. This simple approach is based on the segregation of the image intensity histogram into a predetermined number of classes by threshold values [165, 166]. A threshold value is an intensity value selected, either manually or algorithmically, which groups image pixels into two classes based on their intensity value. Aside from the single threshold method, multiple thresholds or multi-level thresholds can also be derived for an image to better represent real world images in vision applications [167, 168]. An evaluation of different threshold determination techniques is provided in the paper by Boecker et al. [169]. The thresholding method works best in cases where the structure of interest (or colour) falls within a certain range of values which is distinctly separate from its background. This range of values typically apply to intensity images, such as greyscale or components of a colour space. A common way to visualise this range of values is by looking at the histogram of image intensity values.

The threshold for segmentation can be a global or local threshold [170]. A global threshold is a single intensity value selected that divides the entire image into two classes [171], while local thresholds are determined for a local patch in the image [172]. Image thresholding is the simplest method for image segmentation as it only requires searching for a single value that best separates the foreground and background classes. It is used extensively in this research to binarize images based on their intensity values. However, the downside to this method is its susceptibility to noise and its simplicity itself [173]. The performance of the method depends heavily on the correct selection of the threshold value, which is often not a trivial task. Consequently, thresholding is often an image pre-processing operation which leads to more complex operations, such as edge detection and clustering. An example of kidney tubular wall structure segmentation by thresholding is presented in Figure 2.3 (h—i).

2.3.3 Edge detection

Segmentation by the detection of edges in an image is a common method used in cell segmentation [166, 174]. It works by the detection of discontinuities or sudden changes in the image, which usually signals the presence of an object boundary [175]. The discontinuities in the image can be discovered by inspecting the first and second spatial-derivatives of the original image. In other words, the intensity gradient of the image is calculated to look for local edges, which show up as peaks in the intensity gradient. Examples of the gradient-calculating operators are Canny, Sobel, Prewitt, Laplacian and Difference of Gaussian operators [176]. Aside from identifying object boundaries, the edges in an image can also be detected and quantified as image textural features [177].

As with simple thresholding, edge detection is also prone to having its performance affected by noise due to illumination variations [178]. The failure to detect weak edges

or the inclusion of false edges often deteriorates the performance of segmentation. Hence, a smoothing operation is usually carried out before edge detection. A segmentation process based on edge detection needs to be combined with thresholding, morphological or region-based approaches for a complete extraction of the object. In other words, the individual disconnected edges detected are put through a linking procedure to extract the object boundary to be segmented [166]. The edges extracted in the biopsy images of this project are used to outline structure boundaries and as a textural feature for classification, as shown in Figure 2.4.

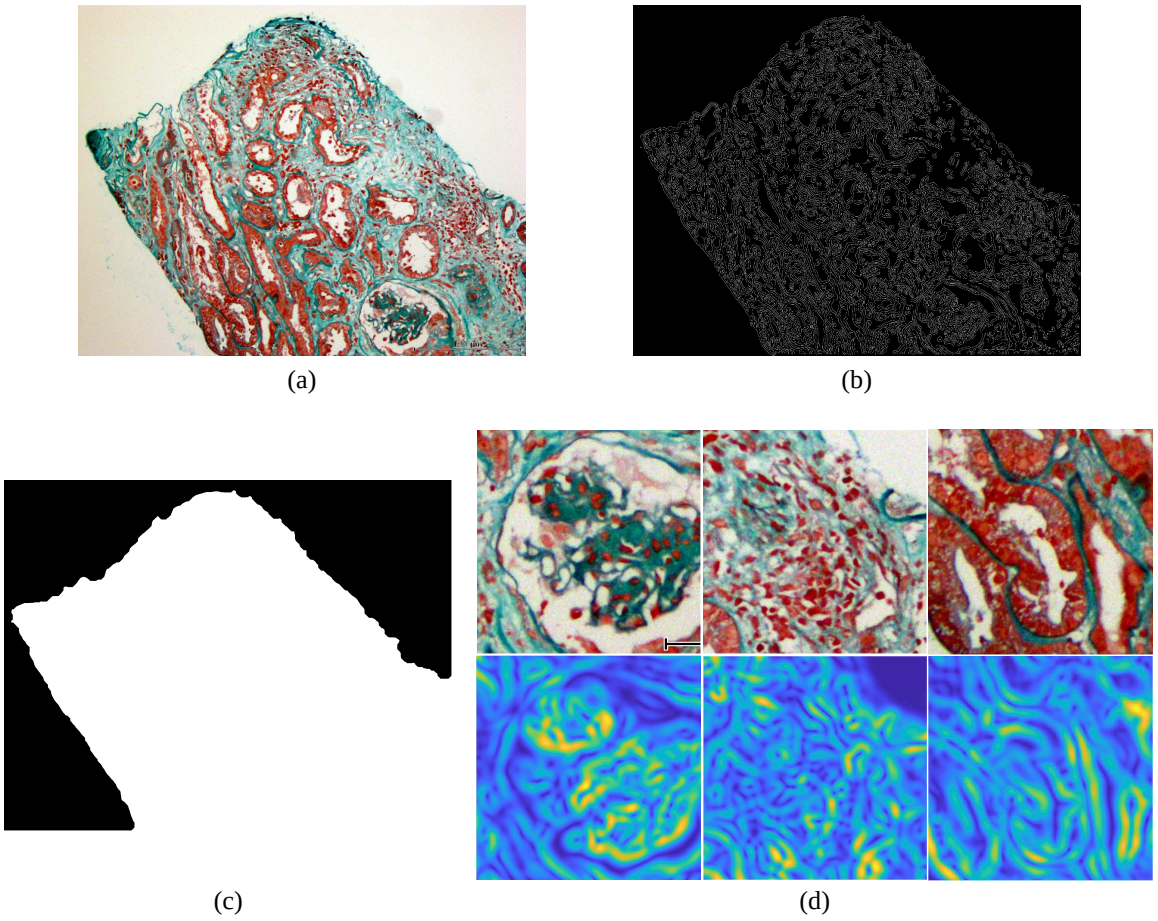


Figure 2.4 An example of the use of image edges and their intensities and orientations on (a) a kidney biopsy image. (b) The intensity edges in the image show different textures for various structures, along with (c) the resulting segmented biopsy tissue. (d) The justification for the use of image edges in the classification of different structures can be seen in the colour edges for the glomerulus, interstitial fibrosis, and renal tubules respectively.

2.3.4 Region processing

A less noise-prone method for segmentation is the region-based approach that groups pixels into homogenous regions based on a selected seed location in the image [179]. This family of techniques remove image-processing noise by incorporating prior assumptions about spatial homogeneity of the pixel values. Examples are region merging, region splitting, and a combination of them [173]. The splitting or merging criterion is commonly the homogeneity of the patch, which is a statistical property and can thus overcome the effect of image speckle noise [180]. The process of the region merging approach starts from the selected seeding points in the image, where surrounding areas are merged if they satisfy the merging criterion. This continues until the entire image is classified. On the other hand, the region splitting approach starts with the entire image as the seed and continuously checks the current patch for homogeneity. The patch splits into four quadrants if it is non-homogeneous and it is repeated until it cannot be split any further. As expected, this method is highly influenced by the selected seed location and the resulting segmentation is constrained by the shape of a square quadrant patch.

Another region-based approach is the watershed transformation, which treats the magnitude of the image as a 3-D topographic image [181]. Each local minimum is treated as either a separate water source or a drainage. The basic working principle of the transformation is either by flooding or draining. For watershed by flooding, the watershed lines or boundaries of objects are defined as the points at which different water sources meet as the water source slowly floods the valleys. In the case of drainage, each pixel membership is related to the downstream steepest decent path towards the closest local minimum. However, watershed transform is prone to over segmentation and usually requires post-processing to refine the segmentation [182].

2.3.5 Textural features

The textural feature of an image is the description of spatial arrangement of the pixel intensities that represents the perceived texture of the image [183]. This description of the pixel intensities is a pattern unique to the image, or a specific section of image corresponding to objects of interest. Texture analysis is even capable of revealing information about the image that might not be obvious in a visual evaluation [184]. Thus, it is a feature commonly used for image classification. Particularly in the case of glomerulus identification, the randomness or complexity of the image regions on the inside and outside of the glomerulus are visually distinct. Thus, this textural information is extracted as a distinguishing feature, as investigated in Chapter 4.

Various textural descriptors can be found in the literature for medical image segmentation. Their applications are typically found in segmentation and classification tasks as a combination set of textures. Examples of image textural features applied in medical image classification tasks include features from GLCM [185], Scale-Invariant Feature Transform (SIFT) [186], multiresolution filters [187–190], Histogram of Oriented Gradients (HOG) [191], and Local Binary Pattern (LBP) [192]. For expositions of these different types of textural features, the reader is referred to the latter section of this subchapter.

Individual textural features have been proven to be useful in classification of medical images due to their ability to rein-in and quantify the randomness of pixel intensity in biological images. Maji et al. [193] and Khuzi et al. [194] have proposed classification methods utilising GLCM features in digital mammogram and brain magnetic resonance images respectively. Different variations of the multiresolution filters have also been explored and developed for distinguishing between healthy and diseased regions of tissue images. Qureshi et al. [195] proposed the Adaptive Discriminant Wavelet Packet Transform for classifying meningioma images, while the Discrete Wavelet Packet Frames

was used by Katouzian et al. [196] for differentiating between blood and plaque signals in intravascular ultrasound images. Similarly, Gabor and Laplacian of Gaussian filters have also been used in extraction of multi-scale image features. For example, Montillo et al. [197] constructed Gabor filter banks to extract the boundary of myocardium in tagged-magnetic resonance images of heart tissues by tuning the parameters of each filter based on approximate location and orientation of tissue boundary. Coppini et al. [198] also used Laplacian of Gaussian filters (multiresolution filters) to enhance salient features in lung radiograms. On the other hand, Alomari et al. [199] employed the entropy of LBP features for breast cancer histology image classification. A variant of the LBP called the circular LBP was proposed as the classifying feature for distinguishing between benign and malignant samples in hyperspectral images of colon biopsy images in a paper by Masood et al. [200].

In more recent studies, it has been increasingly common to combine multiple complementary textural features in order to achieve a higher classification accuracy. For instance, Al-Kadi et al. [201] introduced a set of features encompassing GLCM, Gaussian Markov Random Field, fractals, and run length matrix features for classifying histopathological meningiomas images. SIFT features, which usually consist of many feature data points extracted from key points in the image, are used in combination with greyscale, colour, Sobel, and Tamura histograms in the work by Caicedo et al [202, 203] on basal cell carcinoma images. Similarly, Diaz et al. [204] joined SIFT features with Discrete Cosine Transform coefficients in analysing H&E-stained skin tissue samples. A series of work by Doyle et al. [205–207] integrates the Gabor wavelet features with first order statistical features, GLCM features, and architecture features in their investigation on the classification of prostate and breast biopsy images. The GLCM features were also used in conjunction with Gabor wavelet features in studies by Kalkan et al. [208] and Wang et al. [209] on histology images.

The justification for assimilating multiple textural features into one classification task is that these features may provide complementary information extracted from the image. For instance, the HOG feature is investigated with intensity histogram features in the papers [210–212], while Wan et al. [213] combines HOG features with Kirsch and Gabor filters for a breast cancer histopathology image set. In works by Kumar et al. [214], Rama Krishnan et al. [215], and Mercan et al. [216], the classifying performance of the LBP features are inspected together with deep features, higher order spectra features and colour histogram features respectively. Various types of wavelet transforms were tried and tested in studies done by Lopez et al. [217] and Rama Krishnan et al. [218, 219] with fractal dimension features on prostate biopsy and epithelium biopsy images respectively.

The field of research into content-based image retrieval has also benefited from the advent of more complex image textural features. The content-based image retrieval method extracts descriptive information from large databases of digital images automatically to ease the process of searching for relevant groups of images. Ramamurthy et al. [220] utilises the GLCM features with k-means operation for a medical image retrieval system, while SIFT descriptors were used in Zhang et al. [221] and Cruz-Roa et al. [222] for constructing a visual dictionary of histopathological tissue image datasets. Characteristic image signature generation based on wavelet decomposition coefficients was proposed by Quellec et al. [223] to serve as an image content identifier. Lastly, Zhang et al. [224] have also presented a framework for classification of cell images based on SIFT features with weighted hashing algorithms.

The following subsections expand on the popular image textural features used in computer vision applications.

Grey Level Co-occurrence Matrix (GLCM)

GLCM is a method that examines texture by statistically modelling how often a specific grey-level intensity occurs in conjunction with another grey-level intensity in consideration of their spatial relationship. It calculates the frequency of occurrence for a pair of pixels over a specified spatial relationship with the quantised grey-level intensity of values i and j respectively. This spatial relationship can be between two adjacent pixels or two pixels at a specified offset distance in the horizontal, diagonal or vertical direction. The frequencies of occurrence of these pixel pairs are recorded in a matrix known as the GLCM. From this matrix, various second order statistics can be extracted such as the entropy, contrast, correlation, energy and homogeneity. The GLCM with an offset of one pixel, as an example, would investigate the relationship between a reference pixel and its immediate neighbour on the right. A GLCM tallies the frequency of occurrence for the reference and neighbouring pixel value combination pair. As a simple example, an image with four quantised intensity levels is shown in Figure 2.5.

One of the major characteristics of GLCM is its reliance on a predefined sampling window and the rigid sampling regime. The suitability of GLCM in non-rigid cell structures recognition has not been tested.

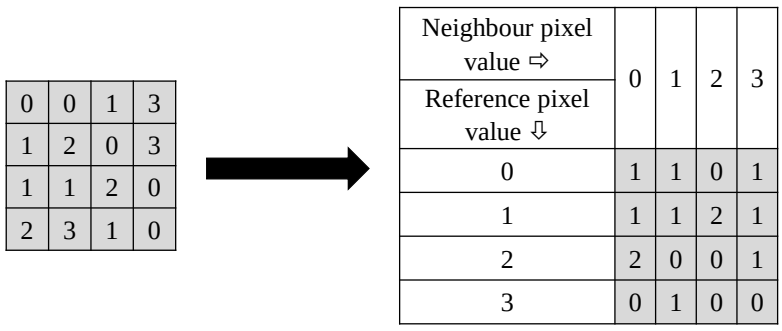


Figure 2.5 Illustration of the GLCM calculation process from a 4×4 image window.

Local Binary Patterns (LBP)

The LBP is an effective texture descriptor involving simple computation and is invariant to greyscale monotonic transformations and image rotation. This method works at the pixel level in a circular manner. Thus, it can reveal local micro-patterns such as edges, flat areas and spots in the texture image. The LBP feature has been proven to be suitable for describing medical images [225].

The LBP texture operator first establishes the circularly symmetric neighbourhood around the centre pixel of interest by defining the radius of neighbourhood in terms of pixel number. This radius is typically small as the correlation between pixels diminishes with distance. Next, the greyscale values of the pixels in the neighbourhood $g_p(p=0,...,P-1)$ are subtracted from that of the centre pixel g_c to find the signs of their differences. By taking the signed differences, the texture descriptor achieves invariance towards shifts in global greyscale levels in the image. Each of the sign $s(g_p-g_c)$ is then assigned a binomial factor 2^p before being summed to produce a unique $LBP_{P,R}$ number for the local texture region. This process is repeated for the whole image to produce a frequency histogram for the occurrence of the different $LBP_{P,R}$ numbers obtained, which becomes the descriptor for a classifier.

For instance, the 3×3 neighbourhood of a pixel can look as in Figure 2.6. Upon comparing against the centre pixel in a clockwise manner, the signed differences would produce a binary matrix which can be converted into a binary number starting from the top left pixel value. This operation is then repeated across the entire image to produce a frequency histogram.

An improved version of the algorithm takes into consideration the much higher occurrence frequency of certain patterns over the others, which are termed 'uniform' patterns. This results in a reduction in size of the $LBP^{ri}_{P,R}$ number set, thereby producing a higher discriminating power from the histogram of the $LBP^{ri}_{P,R}$ number occurrence frequency

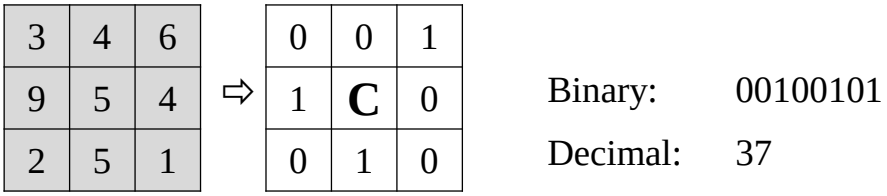


Figure 2.6 The generation of the LBP value from a 3×3 image window.

as a textural feature. LBP has the same limitation as GLCM in the rigidity of the sampling scheme.

Scale-Invariant Feature Transform (SIFT) family

SIFT is a local feature descriptor that is, as its name implies, invariant to scale changes in the image features as well as orientation and illumination changes. This method is based on the extraction of a set of keypoint features in a reference image to be compared and matched to input image features. The SIFT feature vector generated for each keypoint consists of information from the orientation histogram that describes the local region of the keypoint. Mapping the gradient orientations in the context of medical images has shown to be effective for object/structure detection [226]. However, the keypoint mapping of SIFT works better on rigid objects. Non-rigid transformation or variable appearance of cell structures presents a big challenge to SIFT features.

There are four main steps to the SIFT feature generation algorithm. Firstly, the keypoints of the image are determined as local extrema in the different possible scale spaces of the image to attain scale invariance. The scale-space filtering is carried out by a difference of Gaussian operation, which is a type of multiresolution filtering method. The next step pinpoints the exact locations of these keypoints and eliminates the low contrast and false keypoints. The third step attributes the rotational invariance property to each keypoint by assigning an orientation to them based on local image gradient

directions. It is achieved by building an orientation histogram in the neighbourhood of the keypoint and identifying the peaks in the resulting histogram. The final step computes a distinctive keypoint descriptor by creating orientation histograms in a 16×16 neighbourhood around the keypoint to generate a vector. This produces the final 128-element SIFT feature vector. Speeded-Up Robust Features (SURF) is a faster version of SIFT which implements scale-space filtering by box filters of different sizes and makes use of the Hessian matrix to determine points of interest.

Histogram of Oriented Gradients (HOG)

Similar to SIFT features, the HOG method computes the occurrences of gradient orientations in local regions in the image. However, the HOG method calculates the gradient orientation across the entire image instead of certain keypoints by dividing the image into small connected patches termed cells. The gradient in the image is calculated by filtering the image with the $[-1\ 0\ 1]$ and $[-1\ 0\ 1]^T$ kernels. This step produces the horizontal and vertical gradients respectively. This is followed by the creation of cell histograms of the gradient orientations for histogram channels spread between 0 and 180 degrees. Through this step, the dominant gradient orientation in the local region can be determined. To achieve illumination invariance, the histograms are normalised in blocks, which are groups of cells. The final HOG feature vector is then the concatenation of the normalised cell histograms from all the blocks in the image. HOG statistics has been shown to work very well for rigid objects. The statistical model of HOG required to describe non-rigid objects is expected to be complicated.

Multiresolution filters

Inspired by the way the human vision system works, multiresolution texture analysis models are finding reasonably good performance in image classification. One of the

most common method for multiresolution image analysis is the convolution of an image with a bank of filters with tuned parameters to produce multiple versions of the image at different scales. Popular filtering methods include the wavelet transform, the Gabor transform, the Gaussian and Laplacian transforms.

Wavelet transform is carried out by using two different types of filters, namely the low-pass and the high-pass filters. Each filter bank (consisting of a low-pass and a high-pass filter) is then down sampled at a half rate of the previous frequency so as to produce images at lower resolutions. The resulting sub-images are low-low (LL), low-high (LH), high-low (HL) and high-high (HH), corresponding to the filtering on the rows and columns of the original image. Subsequent subsampling would be made on the LL image, as shown in Figure 2.7.

Gabor filters are a type of band-pass filter which is defined mathematically as the product of a Gaussian kernel and a complex sinusoid. The Gabor filter is claimed to mimic the mammalian visual cortex cells and has thus been used extensively in image recognition applications. The Gabor filter bank is typically made up of a set of

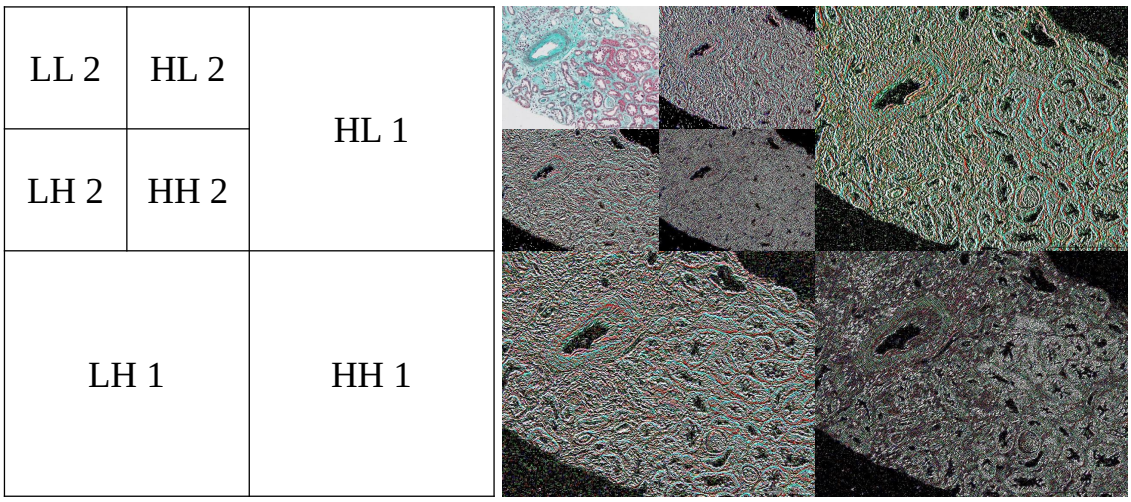


Figure 2.7 Illustration of the wavelet level image decomposition applied on a sample renal biopsy image.

Gabor filters with different orientations and spatial frequencies to explore the multiscale properties in an image.

Similarly, Gaussian and Laplacian filters convolve the image with a Gaussian and Laplace function to produce a version of the image at a different frequency or lower resolution. When the image is iteratively filtered to lower resolutions, the collective images at different scales form a multiresolution pyramid, as illustrated in Figure 2.8. On the other hand, the Laplacian of Gaussian extracts the image texture information by convolving the image with a Gaussian kernel to produce a representation at a specific scale space before applying a Laplacian operator on it. In this case, the Laplacian operator is used to detect maxima and minima in the scale space.

2.3.6 Geometrical features and shape descriptors

The geometrical shape of objects or structures in the medical images is vital information for segmentation. It is typically included as one of the initial steps in the segmentation procedure to determine the approximate shape of the region of interest. Additionally, the morphological statistics obtained from the images is also useful in the classification of tissue sections or the severity of deterioration in tissue samples.

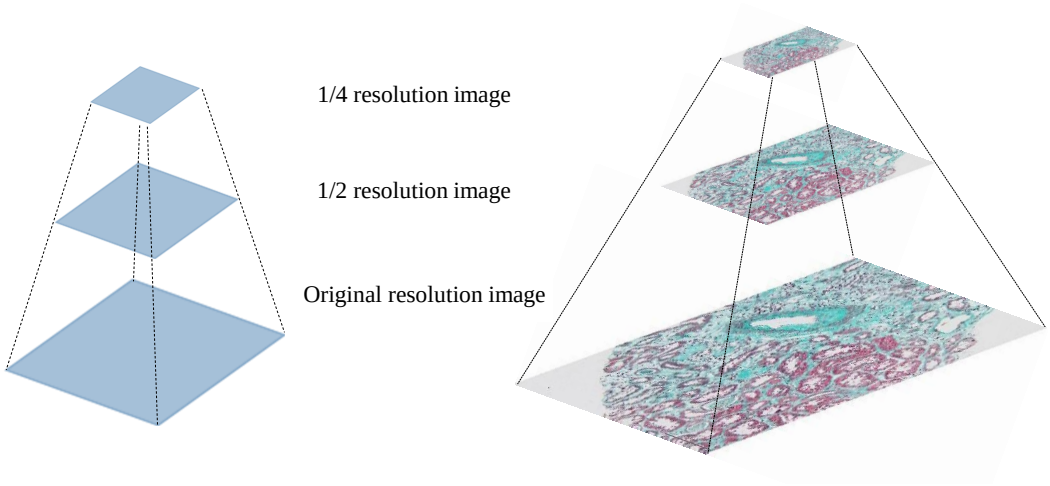


Figure 2.8 An example of the multiresolution pyramid for a renal biopsy image.

Prior to the extraction of geometrical features, the objects of interest (usually cells) has to be segmented by other means, such as thresholding or edge detection. These objects are then inspected for their morphology to be quantified as part of the classification feature. In the papers [227–229], the properties of cell structures were quantified explicitly. Examples of cell morphological features are cell area, height, width, nuclear size, density and shape. Aside from cellular geometry, blood vessels are also a common structure to have their spatial features examined due to their ubiquity in histopathological images. Cooper et al. [230] and Nedzved et al. [231] presented morphological descriptions such as centroid location, size, eccentricity of the objects of interest to describe the characteristics of the tissue sample. Another study by Kwak et al. [232] also obtained geometrical properties of H&E-stained histology images that describe the cell, lumen and general tissue structure to be segmented.

In the context of this research, the size and spatial arrangements of the tissue structures are useful for generalising their characteristics to be subsequently identified and extracted. For example, in the segmentation of renal tubular structures, the thickness and size information of extracted red structures are criteria in the structure validation procedure. However, this feature has to be treated with caution as the geometry of organic cells and structures are inherently highly variable and thus must be complemented by other features.

2.3.7 Graph-based features

Graph-based image features are quantitative descriptors, usually graph-based statistics, that describe the spatial distribution of structures in an image, especially the tissue architecture in medical image datasets. In histopathological images, the density and distribution of the cells or tissue structures provide crucial information to the pathologists regarding the state of tissue health. Graph-based features map the relationship of the

distance between cells of interest in a quantifiable way that facilitates the inference on the architecture makeup of the tissue sample. Hence, graph-based features are also termed architectural features.

The commonly used graph-based image features are the Voronoi Diagram and Delaunay Triangulation. The Voronoi Diagram partitions the data point space into segments based on the bisection of any two points, while the Delaunay Triangulation forms triangles from any three non-collinear points in the data point space. Sharma et al. [233] gives a detailed review on the types of graph-based methods and their use in histopathological image analysis. The architectural features are used in H&E stained breast biopsy tissue images in papers [234, 235] to map the arrangement of cells in the tissue samples. A similar research on breast cancer histopathological images by Basavanhally et al. [236] utilises additional architectural features associated with Delaunay graph in conjunction with a multi-field-of-view classifier to extract features at various field-of-views. Examples of these associated architectural features include minimum spanning tree, nearest neighbour, and relative neighbourhood graph. Studies by Lopez et al. [237] and Guillaud et al. [238] have also used graph-based features in combination with tissue structure morphological features to improve the estimation of cell organisation and distribution in the histopathological images.

A main disadvantage of graph-based techniques is its high computational complexity, which may not be justified for a similar performance by simpler methods of segmentation such as thresholding and geometrical segmentation algorithms. In addition, the number and location of cell substructures change from one sample to another. This poses big challenges to the construction of graph model that is robust and useable across different samples. In the case of renal tissue structures, although graph-based features can potentially provide good descriptions of the tissue architectures, the trade-off lies in the computational power required for the analysis of a high volume of digital image data.

2.4 Classification techniques

With the extraction of identifying features performed on the image of interest, the classification step can be carried out using several types of machine learning algorithms. The type of classifier chosen is typically based on the operational requirements and constrains imposed by the application. Table 2.1 gives a brief overview on the common classifier types used in CAD systems.

Of the classifiers introduced here, only decision tree, SVM, and ANN (specifically convolutional neural network) are used in this investigation. Decision trees are able to provide insights to the decision-making process for classification, which is valuable information in medical diagnoses. This is contrasted against the ANN, which is able to achieve competitive classification performance but unable to offer a clear reasoning to its classification. LDA is a special case of ADTree to be introduced in Table 2.1.

Table 2.1 A summary of various classifiers employed in medical imaging and diagnosis applications.

Classifier	Description
Linear Discriminant Analysis (LDA) [239]	A supervised approach for classification that finds a linear combination of features that separates two or more classes of objects or events. It is a traditional statistical classification method that is derived through the optimality principle. The availability of efficient numerical linear algebra packages made the implementation extremely efficient and simple. It works very well on a class of data that is termed linearly separable. However, LDA is less successful in non-linear dataset with complex topology.

Decision tree [240]	A graphical model that splits the training data into branches until it ends in a terminal node, which contains a final class prediction. The splitting heuristic is based on thresholding of individual or multiple features towards achieving homogeneity in the class label. This method provides a comprehensible visualisation of the decision rules made in achieving the classification, but the accuracy of the model depends heavily on the design of the tree. This is because the reliability of the decision tree relies on the optimal parameters fed into it.
Sparse Alternating Decision tree (Sparse ADTree) [241]	Sparse ADTree is a classifier that generates a classification model of the training data by simultaneously producing a sparse feature vector using Elastic Net regularisation [241]. In other words, the weight corresponding to each element in the feature vector is updated to reflect its importance in the modelling of the classification model [242]. Since the Sparse ADTree has been proven to achieve high performance with highly correlated features, it is the chosen method for feature selection and classification in this study. One characteristic feature of the Elastic Net regularisation is that it allows the weight of an element to be zero. Therefore, these feature elements with zero coefficient are effectively removed from the formulation of the classification model's decision nodes.

The Sparse ADTree employs the Sparse Linear Discriminant Analysis for the purpose of performing feature selection and classification concurrently. The LDA algorithm employs optimal scoring to cast the classification task as a regression problem by turning the class label from a categorical element into a quantitative element, $\boldsymbol{Y}$. The algorithm iteratively minimises the residual or least squares difference between the $\boldsymbol{Y\theta}$ and $\boldsymbol{X\beta}$ (where $\boldsymbol{X}$ is the dataset matrix, $\boldsymbol{\theta}$ is the optimal score vector, and $\boldsymbol{\beta}$ is the discriminant vector) in search of a linear regression model that best approximates the desired output.

$$minimise \left\| \boldsymbol{Y\Theta} - \boldsymbol{X\beta} \right\|^2 \tag{2.1}$$

The Elastic Net regularisation is introduced to the LDA algorithm to induce sparsity in the feature space by the addition of the Lasso penalty term $\lambda_1 \left\| \boldsymbol{\beta} \right\|^2$ and the Ridge penalty term $\lambda_2 \left\| \boldsymbol{\beta} \right\|_{\mathbf{1}}$. The optimal score vector $\boldsymbol{\theta}$ and discriminant vector $\boldsymbol{\beta}$ are optimised through an iterative algorithm of alternating between fixing one parameter while solving for the other parameter until convergence.

$$minimise \left\| \boldsymbol{Y\Theta} - \boldsymbol{X\beta} \right\|^2 + \lambda_1 \left\| \boldsymbol{\beta} \right\|^2 + \lambda_2 \left\| \boldsymbol{\beta} \right\|_{\mathbf{1}} \tag{2.2}$$

	The LARSEN (Least Angle Regression for Elastic Net) algorithm used in solving for the $\boldsymbol{\beta}$ in the iteration produces a series of $\boldsymbol{\beta}$ solutions for different numbers of features, starting from zero to the entire feature vector length, which is also termed the regularisation path. This allows the optimal sparsity of the feature set to be determined based on the model selection criteria. In the case of the Sparse ADTree, the cross-validation generalisation was the chosen model selection criterion. The classification features are assessed in terms of the sparse feature length and their respective classification accuracy.
Regularised Logistic ADTree (rLADTree) [243]	Similar to the Sparse ADTree, the rLADTree is a multivariate ADTree which performs learning by incorporating boosting cycles in its algorithm. However, the rLADTree induces a multivariate ADTree through the use of the boosting technique called LogitBoost instead of sparse LDA. LogitBoost is a boosting algorithm that utilises a linear combination of regression models to produce an estimate of the log-odds ratio between positive and negative class posteriors. A weighted least squares regression problem is solved in each boosting cycle of LogitBoost, where the regression model used is limited to be of a linear type. At the end of every boosting cycle, the regression values for all training samples are updated to produce a new probability estimation for each sample. Classification using the rLADTree is carried out by taking the sign of the summation of the resulting regression function.

k-Nearest Neighbour (kNN) [244]	An example-based learning method that assigns the output class of an object to the most common class among its k nearest neighbours. It is one of the simplest machine learning algorithm and requires a good choice of the k value in order to reduce the adverse effect introduced by noisy data. However, the downside of this method lies in the fact that the classification is based purely on existing instances of data, therefore the presence of noise or irrelevant features can severely degrade the quality of kNN classification.
Support Vector Machine (SVM) [245]	A classification method that maps the features into a high dimensional space to search for a hyperplane that optimally separates the training datapoints into the corresponding two categories. SVM can achieve a good repeatable classification result similar to LDA for linear datasets. On non-linear datasets, the "kernel trick" can be used to implicitly map the input onto high-dimensional feature spaces, converting nonlinear problem into a linear one. The prerequisite to the success of this supervised learning method is the availability of labelled training data and the setting up of optimal kernel parameters. These requirements are sometimes costly and time consuming, which is one of the disadvantages of the SVM.

The most popular SVM kernel functions used are the linear, polynomial, and Gaussian (or radial basis function) kernels. Given an identically independent distributed training set $\{x_i, y_i\}$, where $x_i \in \mathbb{R}^d, y_i \in \{-1, 1\}, i = 1, ..., n$. In the case of a non-linear dataset, the kernel function maps the input samples onto a feature space to obtain a linearly separable training set. The problem can be converted to the maximisation of a dual optimisation problem,$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j K(x_i, y_i)$, subject to $\sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \in [0, C]$ for $i \in [1, n]$. Here, the $K(x_i, y_i)$ represents the kernel function.

The frequently used SVM kernels are:

1. Linear kernel: $K(x_i, y_i) = (x_i \cdot y_i)$
2. Polynomial kernel: $K(x_i, y_i) = (x_i \cdot y_i + 1)^p$, where p is the degree of the polynomial
3. Gaussian (radial basis function) kernel: $K(x_i, y_i) = exp(-\frac{\|x_i - y_i\|^2}{2\sigma^2})$

Artificial Neural Network (ANN) [246]	Inspired by biological neural networks, this algorithm is a collection of interconnected units called nodes that computes their output as a non-linear function of the sum of its inputs. These nodes change their parameters or weights dynamically as the information passes from the input to the output. This method is robust and requires no labelled data. However, it is unrepeatable and demands a long training time. ANN does not lead to transparent decision-making rules in stark contrast to other algorithms mentioned above which make use of mathematical abstractions.

2.5 Other Methods

Some methods reported in the literature do not strictly conform to the workflow shown in Figure 2.2. Most of these methods do not have clear feature generation and classification stages. The features are either not needed or will be generated internally as a side-product of classification. This section will briefly review the most popular methods among these alternatives.

2.5.1 Clustering

Clustering works by grouping similar image pixels or data points by a certain set of similarity criteria in order to identify objects of interest. These similarity criteria are defined in terms of the distance between points in the feature space of interest, such as the greyscale intensity or colour space component intensities of an image [247]. This method is an unsupervised segmentation task as it attempts to classify unlabelled data

into clusters through a number of iterations [248]. In each iteration, the algorithm clusters the data in a manner which maximises the decision boundary between the different groups. The performance of clustering algorithms is mainly dictated by the goodness of the initialising parameters and the stopping criterion [249]. It is at risk of falsely converging onto a local minimum, depending on the initialising values and distribution of data. Additionally, this method also requires a predetermined knowledge on the number of cluster, which might be unknown in most cases [250]. However, an advantage of clustering algorithms is its high computational speed [251]. Hence, depending on the application and robustness required of the task, different variants of the clustering algorithms can be selected from to suit the need. The different variants of clustering algorithms include k-means, C-means, fuzzy C-means, and EM algorithm [252]. The k-means is chosen in this thesis for its simplicity and robustness to find homogeneous regions in the images.

2.5.2 Deformable model fitting

Deformable models are template shapes or contours that are gradually distorted and fitted towards outlining the boundaries or edges of an object of interest [253]. The template shapes used are the approximate outlines of the object to be segmented. This fitting operation is based on the minimisation of energy in the shape model [254]. In other words, there are two opposing forces at work which slowly moulds the shape into the actual object outline [255]. These two forces are the internal energy which tries to retain the smoothness of the original template shape and resist changes to the original shape, while the external image energy drives the shape towards the object boundary. The introduction of a closed parametric curve or shape as an initialising shape allows the direct segmentation of the object outline with minimal effect from any image noise. Its performance is, however, still subject to the accuracy of the initial location and shape of the model chosen by the user [256]. Since the kidney tissue

structures are subject to high variation in their appearances (the largely varying shapes of renal tubules in Figure 2.4(a) of Chapter 2.3.3, for example), a deformable model was not selected to be included in the segmentation algorithm.

2.5.3 Markov random field models

The Markov random field is a type of undirected graphical model capable of modelling the relationship between a pixel and its neighbouring pixels [257]. By learning the spatial correlation between pixels, it produces a statistically rigorous model that is incredible robust against imaging noise. The Markov random field specifies the local characteristics of the image by modelling the probability of a pixel values based on its neighbourhood [258]. For an image segmentation task, the process is carried out by finding the *maximum a posteriori* estimation of the expected pixel values [259]. This would result in a prediction model of the segmentation class for the target pixel. An advantage of the Markov random field lies in the fact that it can adopt constraints on the system modelling to account for interactions between pixels in the image by integrating an *a priori* probability into its equation. However, the parameter estimations for the strength of spatial interactions between pixels has to be carefully chosen for maximising performance [260]. In a complex image like the kidney biopsy image which consists of large amounts of irregularities, the Markov random field model required may be too complex to fine tune.

2.5.4 Deep neural networks

The older generation of ANN classifier generally follow the workflow shown in Figure 2.2. However, there are recent developments of ANN that follow a completely different workflow and were given the name of Deep Neural Network. The ANN approach tackles the segmentation problem by deriving general classification rules using multiple

networks of processing nodes that work in parallel [261]. A deep neural network requires a typically large training image dataset and demands a long training time [252]. However, once trained, the classification process is relatively quick. Deep neural networks model the target system by performing transformations over several layers of interconnected neurons (computation nodes) from the input to the output, each node with its own assigned weight. The interconnectivity of nodes between the layers allows high flexibility in the modelling of non-linear systems, which is an attractive trait for classification tasks [262]. In the domain of image segmentation, neural networks can be used to classify pixels into object groups based on the training image dataset provided [263, 264].

One of the most popular deep neural network architecture is the convolutional neural network (CNN). While basic ANNs have fully connected layers, CNN uses convolutional kernels and pooling layers to reduce the number of parameters, thus speeding up the learning process. The architecture of CNN is very similar to the signalling pathway of mammalian visual cortex, hence many believe that CNN has sufficient information processing capability to solve most visual recognition problems. The most important prerequisite to a good generalisation by CNN is the availability of a large-scale labelled dataset to avoid over-fitting [265]. In the medical setting, the main difficulty is presented not only in the availability of the image data, but also in the acquisition of the relevant annotations for these images by experts. Thus, in small datasets with limited amount of training data, handcrafted features are often the more ideal choice for use in classification [266]. Furthermore, it has also been revealed that the distinguishing criterion for training a successful neural network is not in the architecture design itself, but the expert opinion on the type of information or data to feed into the network (such as an additional pre-processing stage or data augmentation) [267]. In other words, the success of CNN still rests heavily on the explication of expert knowledge, and a high performing CNN does not necessarily provide insight to how the data is generalised. Hence, this thesis

investigates the suitability of CNN features in the classification task of glomerulus structure identification in contrast with other handcrafted features.

2.6 Glomerulus segmentation methods

The glomerulus is a distinguishing structure in the kidney responsible for the filtering of excess water and metabolic waste products from the blood. Pathologists examine the structure of the glomerulus as well as other disease indicators, such as IF, when examining a kidney biopsy sample as an indicator to the kidney's state of health. Glomerulus structure segmentation is an important process in the computer-aided diagnosis of the amount of IF present in the biopsy sample [268]. The automatic identification and segmentation of glomerulus structures in biopsy images remain a challenging task in the literature due largely to the structure's morphological changes in diseased states and its varying appearance depending on the biopsy slice orientation. The glomerulus structure and its segmentation method are described in Chapter 4.1. The reader is referred to Figure 4.2 for an introduction to the structure's appearance.

Yamada et al. [269] proposed a dynamic programming application to detect the glomerulus locations in kidney microscopic images. Dynamic programming is a model matching method which requires a predetermined polygonal model for the contour of the glomerulus. Similarly, the paper by Maree et al. [137] employs the bioimage analysis platform named Icy which also performs detection based on the elliptical-shape assumption of the glomeruli. Building on the shape assumption, the paper by Zhao et al. [270] has specified three empirically-determined descriptors for the typical Bowman's capsule. These morphological descriptors are the aspect ratio, solidity, and circularity of candidate objects matching the characteristic colour of the glomerulus after biopsy staining. The work by Kotyk et al. [271] employs the same morphological descriptors for glomerulus detection while a convex hull fitting step was introduced to refine and

establish the boundary of the Bowman's capsule. The performance of these methods, with their focus on the contour of the glomerulus as the basis of detection, are heavily dependent on the goodness of appearance of the Bowman's capsule. The elliptical assumption on the shape of the glomerulus has been tested and held true in much of the research work in the literature. Therefore, this work also incorporates this assumption into the segmentation of the glomerulus structure.

Zhang et al. [272, 273] attempted to overcome the staining variation problem by detecting the edges of the Bowman's capsule instead of its colour. Genetic algorithm was proposed to search for the optimal spline-fitting curve that defines the glomerulus boundary. Further improvements of the algorithm were also proposed by incorporating the watershed transform [274], wavelet transform [275], and the self-organising feature map neural network [276]. The purpose of the self-organising feature map is to perform colour clustering on the biopsy image for the segregation of the region of interest from the background. On the other hand, Pedraza et al. [277] achieved a high accuracy performance by using a pre-trained CNN model on the task of classification between glomerulus and non-glomerulus images.

Meanwhile, Kamenetsky et al. [278] proposed a split and merge method with a check on the grey level homogeneity of the local region to segment the glomerular basement membrane area. The work assumes that the basement membrane region is a relatively homogeneous region compared to its background. Ginley et al. [279] utilised Gabor filtering on a Gaussian blurred greyscale image of the glomerular tissue image to approximate the location of the glomerulus. The boundary of the glomerulus is then refined by combining statistical F-testing and a distance transform from the centre of the image. On the other hand, the HOG was introduced as the feature used to train the glomerulus detection classifier in the paper by Hirohashi et al. [280], while Kato et al. [281] proposed an improved version called the Segmental HOG descriptor which is based on the computation of boundary likeliness on the estimated glomerulus boundary.

In contrast, a combination of features including the HOG, 3D colour histograms, means and standard deviations of the RGB colour channels, and mean grey values of the image was presented by Gadermayr et al. [282] as the classifier training data. Inspired by these approaches, a new feature descriptor for image classification is proposed in this thesis to be applied in glomerulus detection.

2.7 Existing interstitial fibrosis quantification systems

Computer-aided diagnosis of fibrosis is not only possible for the kidney. Automatic quantification of fibrosis in other organs requires the consideration of structures specific to the organ in the development of the computer-assisted quantification framework. In the inspection of myocardial biopsy samples, Daunoravicius et al. [283] proposed a colocalization algorithm on MT-stained samples to distinguish features for training a fibrosis-detecting classifier. Also, an ImageJ macro software was developed by Hadi et al. [136] to quantify myocardial fibrosis in Picosirius red-stained slides using a pre-set colour threshold. These systems employ stereology to estimate the three-dimensional geometry of the heart. Besides the heart, fibrosis quantification is also relevant in monitoring the presence and progression of lung diseases. Marten et al. [284] presented a quantification software based on a region-growing method and watershed transform. Taylor et al. [285] utilised Laser Scanning Confocal Microscopy to obtain an intensity image of the lung biopsy that can be subsequently thresholded for extracting fibrotic region.

However, perhaps the most extensive research into automated fibrosis quantification by image analysis, with the most number of commercially available products, is done on liver biopsy images. In [286], Masseroli et al. proposed an application named FibroQuant which performs a very similar fibrotic area extraction method to that used in [287]. The difference is in the segmentation of structures specific to the liver,

such as the portal-periportal and septal area, which are extracted interactively by the user. Another proposed quantification system for liver fibrosis is the CellProfiler by Sant'Anna et al. [288]. The system segments the Goldner's MT-stained fibrotic region by performing operations on the *RGB* colour components of the digital biopsy image. A user input of the typical diameter of the object to be segmented is required for the system to gauge the scale of the digital images. Similarly, Meejaroen et al. [289] have also exploited the colour information presented by the trichrome stain by creating a Bayesian classifier to segment fibrosis pixels in *RGB* liver biopsy images. Unfortunately, the algorithms for fibrosis quantification in other organs cannot be directly used for kidney tissue analysis as their tissue structures are vastly different. Hence, a separate set of rules and techniques are developed for kidney biopsy sample inspection.

2.8 Renal interstitial fibrosis quantification systems

In general, there exist several proposed methods for automated computerised image analysis in quantifying fibrosis in a biopsy specimen. Much of the research in this area is based on the premise of point counting of fibrotic region in stained biopsy samples. In other words, the total fibrotic area is calculated as a percentage of the whole biopsy tissue sample. For example, Masseroli et al. [286] performed automatic extraction and quantification of renal fibrosis area on Sirius red-stained biopsy samples by using an automatic threshold and morphologic filtering method to detect the fibrotic region. The stained fibrosis area was segmented by a global thresholding method from Kurita et al. [290] to produce a binary image for subsequent classification of features based on their shape and size. Similarly, Servais et al. [291] also presented an automatic quantification method that extracts the green-coloured region of trichrome-stained renal biopsies, which is the colour indicative of IF for the stain. On the other hand, Moreso et al. [292] employed a texture analysis method instead of attempting to segment tubulointerstitial

structures in quantifying renal damage. This decision stems from the claim that the similarity between the brightness of the features to be extracted and the background will cause a loss of information in segmentation.

Yet, a higher accuracy of IF quantification cannot be achieved without positive identification of the segmented structures or regions of interest. This approach has been adopted by Klapczynski et al. [293] in rat renal tissue specimens, while a more comprehensive system for identifying human kidney structures was presented by Meas-Yedid et al. [15]. The physiology of the rat renal tissue is a good model for the study of human kidney tissues, but it is still no substitute for actual human renal tissues. The quantification system presented by Meas-Yedid et al. proposed to segment the green colour of the trichrome staining in renal biopsy images by performing colour clustering in a self-defined *I1H2H3* colour space. Since the IF structure is not the sole structure stained green by trichrome stain, the system proceeded to remove other non-IF structures such as the renal capsule, basement membrane, glomeruli, and blood vessels. However, there are three main shortfalls in the system. The first being its segmentation based on the premise of green pixel detection is highly subject to stain quality and sample age, as the green colour contrast fades significantly with age compared to the red colour. Secondly, pathologist feedback suggested that the quantification for interstitial fibrotic area is not limited to only blue-green regions, following certain assumptions in the tissue structure. The justification for these assumptions is introduced and explored in Chapter 1.1. Lastly, the system does not incorporate the segmentation of healthy glomeruli, of which signs of kidney failure can also manifest in the form of degradation in the glomerulus structure. This degradation, termed glomerulosclerosis, is routinely checked for by pathologists in their diagnosis for chronic kidney diseases. Segmentation of the glomerulus structure is thus a useful additional information to be extracted from renal biopsies. Therefore, this research proposes a comprehensive computerised quantification framework for renal IF in stained biopsy images that addresses these shortcomings.

Chapter 3

Knowledge-based Renal Interstitial Fibrosis Quantification

This chapter explores and defines the anatomical information obtainable from the biopsy tissue image dataset used in this research. The focus is placed on the selection of the optimum method for IF quantification based on a literature review of the relevant field of medical image segmentation (Chapter 2). A majority of the research employed a structure segmentation algorithm based on the colour information extracted from chemically-stained biopsy images. Therefore, an investigation into the various colour spaces was conducted in an effort to maximise the amount of information extractable from the colour biopsy images. The information derived from the biopsy image in turn allowed the kidney tissue structures present in the biopsy to be inferred, and subsequently segmented for quantification.

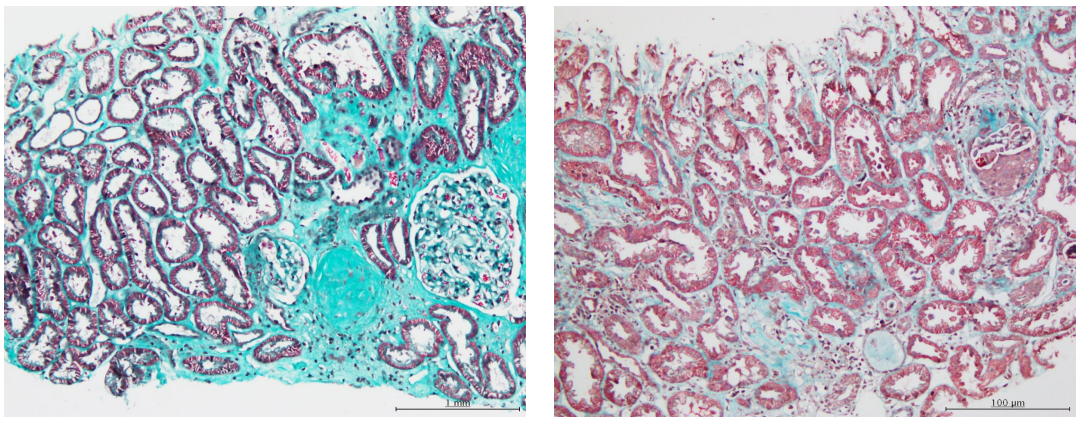
Furnished with the results from the preliminary investigation on the best approach for tissue structure segmentation, the detection and extraction steps for each of the kidney tissue structures were then derived individually. The accuracy of the quantification was statistically analysed and compared against the image dataset with annotated ground

truth. Its performance was also compared against the quantification performed by a group of trained pathologists in order to provide a direct correlation to the actual practice in the diagnostic process.

3.1 An investigation into colour segmentation

As can be observed from Figure 3.1, the MT-stained biopsy slides are clearly presented in three colours, namely blue-green, red and white. The tissue's collagen structures (fibrosis) appear blue-green, the keratin and muscle fibres (tubules and artery walls) are stained red, while the cellular fluid spaces appear white. A simple and well-explored method of quantifying the fibrosis region is by counting the number of blue-green-coloured pixels in the biopsy tissue. Papers [6, 136, 288] are examples of studies exploiting the colour information in the biopsy images for segmentation of fibrosis.

However, it has been observed from the digital biopsy image acquisition process that the variation in MT stain concentration and the age of the biopsy sample influence the contrast between the three colours in the stained biopsy sample. This is especially



(a) (b)

Figure 3.1 A comparison of the colour contrast of the stained specimen in (a) a freshly stained biopsy sample, and (b) an older biopsy sample. Notice the loss of contrast, especially between the blue-green stained region and the white background.

evident in the fading of the blue-green colour into the white background, causing a loss of information from the image, as demonstrated in Figure 3.1. In an attempt to recover the colour information loss, an investigation on the effect of several colour space transformations is conducted, with the algorithm shown in Figure 3.2.

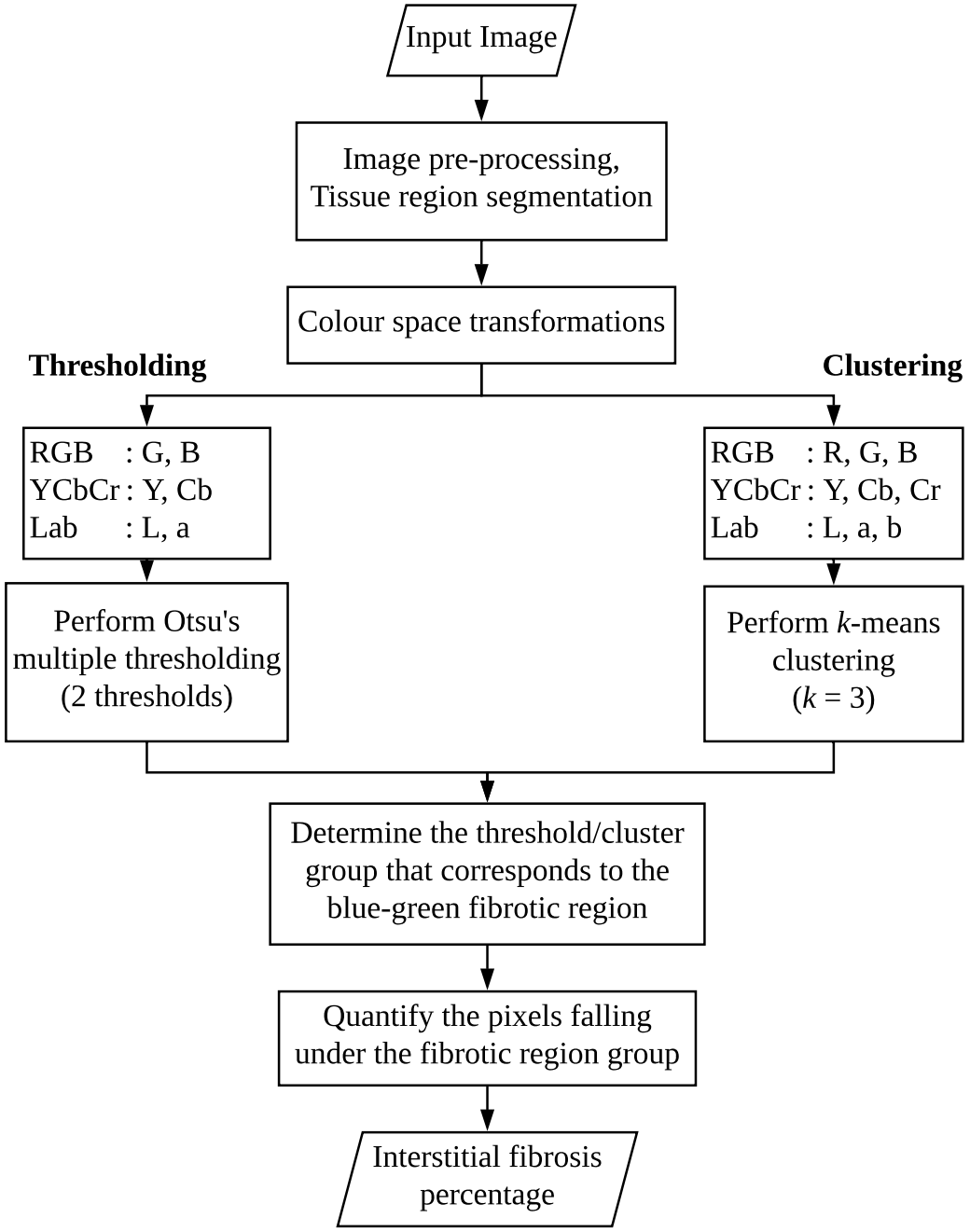


Figure 3.2 The algorithm for the investigation of the effect of different colour spaces on IF segmentation by colour information.

The colour images of biopsy samples are conventionally processed in the *RGB* colour space. However, in computer vision applications, a colour image can be represented in several different colour spaces [294]. A colour space is effectively the description of the range of colours that can be reproduced in an image. Depending on the application, digital images can be transformed between different colour spaces to maximise its efficiency in the later stages of processing. A major weakness of the *RGB* colour model is its inability to separate the luminance (or brightness) and chrominance (colour) components, making it a poor choice for colour analysis and colour-based recognition [141].

To overcome this weakness, the use of the *YCbCr*, *CIE Lab*, and *HSI* colour spaces is investigated in this study. These colour spaces interpret colour by separating the chroma component (colour) from the luma component (brightness) in an image, thus providing the means to identify the difference in colours present in the image in a more consistent manner. Table 3.1 summarises the brightness and colour components of these colour spaces.

The *YCbCr* colour space is a colour image encoding that allows for reduced bandwidth in chrominance components compared to the luminance component. This is justified by the fact that the human vision is significantly more sensitive to changes in brightness rather than colour. It is widely used in video and digital photography systems. The transformation from the *RGB* to *YCbCr* colour space is given by Equations (3.1—3.3) [295].

The *CIE Lab* colour space was proposed by the Commission Internationale de l'Eclairage as a colour-opponent space that is designed to approximate human vision. It includes all colours of the *RGB* colour model, exceeding even those perceivable by human. The transformation from the *RGB* to the *CIE XYZ* colour coordinates is expressed in Equation (3.4) [296]. Dimensions for the *Lab* colour space include lightness L^* and two colour-opponent dimensions a^* and b^* that represent the red/green and the yellow/blue

opponent colours respectively, as derived in Equations (3.5—3.7) [296]. *Lab* colour space is device independent and covers an infinite possible subset of colour representations. Thus it is often used as an interchange format between different devices using different colour spaces.

The *HSI* colour space models the *RGB* colour space more intuitively using a cylindrical-coordinate system, as presented in Equations (3.8—3.11) [176]. Information corresponding to the colour aspect of the image is expressed by the hue (H) and saturation (S) values. They form the circular surface of the cylinder while the intensity (I) is represented by the height of the cylinder. H describes the basic colour that appears in the image (red, blue, purple, etc.) while S represents the degree of difference between colour and grey.

Table 3.1 The brightness and chrominance components of the colour spaces investigated.

Colour Space	Transformation Equations from *RGB* Colour Space
YCbCr	$$Y' = 16 + \frac{65.739 \times R}{256} + \frac{129.057 \times G}{256} + \frac{25.064 \times B}{256} \quad (3.1)$$
Luma Component: Y'	$$C_b = 128 - \frac{(37.945 \times R)}{256} - \frac{(74.494 \times G)}{256} + \frac{(112.439 \times B)}{256} \quad (3.2)$$
Chroma Components: Cb, Cr	$$C_r = 128 + \frac{(112.439 \times R)}{256} - \frac{(94.154 \times G)}{256} - \frac{(18.285 \times B)}{256} \quad (3.3)$$

CIE Lab

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{3.4}$$

Luma

Component: L^*

$$L^* = 116 \left(\sqrt[3]{\frac{Y}{Y_o}} \right) - 16 \tag{3.5}$$

Chroma

Components:

a^*,b^*

$$a^* = 500 \left(\sqrt[3]{\frac{X}{X_o}} - \sqrt[3]{\frac{Y}{Y_o}} \right) \tag{3.6}$$

$$b^* = 200 \left(\sqrt[3]{\frac{Y}{Y_o}} - \sqrt[3]{\frac{Z}{Z_o}} \right) \tag{3.7}$$

HSI

$$H = \begin{cases} \theta & for\, B \leq G \\ 360^\circ - \theta & for\, B > G \end{cases} \tag{3.8}$$

Luma

Component: I

$$\theta = \cos^{-1} \left\{ \frac{0.5[(R-G)+(R-B)]}{\sqrt{([(R-G)^2+(R-B)(G-B)])}} \right\} \tag{3.9}$$

Chroma

Components:

H,S

$$S = 1 - \frac{3}{R+G+B} min(R,G,B) \tag{3.10}$$

$$I = \frac{R+G+B}{3} \tag{3.11}$$

The two methods compared in this part of the research for the extraction of fibrotic regions are colour segmentation by thresholding and by colour clustering. For the thresholding method, the colour space components involved are those that correlate to the blue-green colour indicative of the fibrotic structure as stained by MT, namely the G and B of the RGB colour space, Y' and Cb of the $YCbCr$ colour space, and L^* and a^* of the *CIE Lab* colour space. For each of the colour space components, the Otsu's method [297] is employed to dynamically determine the suitable threshold value in order to segregate the image pixels into three distinct groups reflecting the three colours in the image. This allows the identification of the pixels of fibrotic regions in the biopsy image. Two threshold values,$t_{1,colour}$ and $t_{2,colour}$ (where $t_{1,colour} < t_{2,colour}$) are assigned to each of the colour space component to classify each image pixel into one of the three groups based on their colour pixel values in the corresponding colour space. The threshold group that corresponds to the colour of fibrosis structures is determined based on the criteria specified in Table 3.2.

On the other hand, the colour segmentation based on colour clustering utilizes the k-means clustering method [298] to identify blue-green coloured pixels which represent the fibrotic region. The main limitations of k-means clustering, when compared to other clustering algorithms such as mean shift clustering [299], is the selection of the number of means, k. This number has to be carefully selected before the algorithm is run to ensure the data is optimally partitioned. In this instance, the k value is set to three to reflect the three main colours in the biopsy images. The selection of this k value is justified by referring to the three principal colours (white empty spaces, red keratin or

muscle structures, and blue-green fibrosis or collagen tissues) presented by a MT-stained biopsy sample, as distinctly illustrated in Figure 1.1. Another point of concern is the randomly selected initial means, as the k-means algorithm may converge to different points in the data space based on the presence of local minimums. To minimise this effect, the initial means are experimentally determined based on the average values of each of the three main colours present in the MT-stained biopsy images. Upon successful clustering of the colour pixel values into three clusters, the fibrotic region is identified by the cluster mean value. Fibrosis pixels are those grouped to the cluster with the second highest value for the luma component (Y' and L^* channels for *YCbCr* and *CIE Lab* colour spaces respectively) or the second highest value for the B channel of the *RGB* colour space.

Table 3.2 summarises the conditions for classifying image pixels in each of the colour space investigated for both thresholding and clustering methods. Upon the identification of pixels corresponding to the colour of fibrosis structure, the amount of these pixels is calculated and expressed as a percentage of the total number of pixels of the biopsy tissue region. The experiment comparing the accuracy of these two methods is explained in Chapter 3.3.

Table 3.2 The criteria for classifying an image pixel as fibrosis structure in each colour space for the colour thresholding method and colour clustering method.

Colour space	**Fibrosis classifying criteria in each colour space**	
	Thresholding	**Clustering**
RGB	$t_{1,G} < G < t_{2,G}$ **AND** $t_{1,B} < B < t_{2,B}$	$B == B_2$
YCbCr	$Cb > t_{2,Cb}$, excluding $Y' > t_{2,Y'}$	$Y' == Y'_2$
Lab	$a* < t_{1,a}$, excluding $L^* > t_{2,L}$	$L^* == L^*_2$

3.2 Tissue structure identification and segmentation

3.2.1 Tubular structure segmentation

As noted in the previous section, the method of identifying the IF structure by its blue-green colour presents a difficulty in poorly stained biopsy samples due to the diminished contrast between the fibrosis colour and the white background. Hence, it is proposed to employ a method of deduction in the identification of IF by expanding the segmentation to include all structures of the kidney biopsy tissue. By extracting all non-IF structures from the biopsy tissue area, the leftover areas must be the IF structure. This decision stems from the observation that the contrast of red colour in the biopsy images remained high in the range of biopsy images with staining and age variations. Additionally, five out of eight of the structures in Table 1.1 appear in red. Among these structures, the inflammatory cells are included in the quantification of renal IF. Without the segmentation of all kidney tissue structures, the inflammatory cell structures would be excluded and cause an inaccuracy in the quantification of IF. Therefore, using just colour information may present a false target error in the result. Additionally, discrepancies in the areas highlighted by the stains due to chemical variations in the stain, human error and slide contamination are common.

The main structure present in the biopsy slides is the tubular structure. Tubule walls are stained red by MT, and typically presented as a ring of red walls enclosing blank, white region. The active shape model [300] is a useful statistical model for shapes of objects. It iteratively deforms a general shape of the object to fit an instance of the same object in another image. However, this method requires an initial assumption on the shape of the object to be identified. Since the tubules can appear as a wide range of shape and size, depending on the orientation at which the tubule section is cut, the active shape model will not work efficiently on tubule segmentation.

The proposed algorithm for the segmentation of tubular structures employs a hybrid-colour space red pixel detector, and its flow is presented in Figure 3.3. This method combines the effects of different colour space components that correlate with the intensity of the colour red in the image. These include the *Cr* component of *YCbCr* and the *a** component of *Lab*. In addition, since the *Cb* component is a measure of the colour blue,

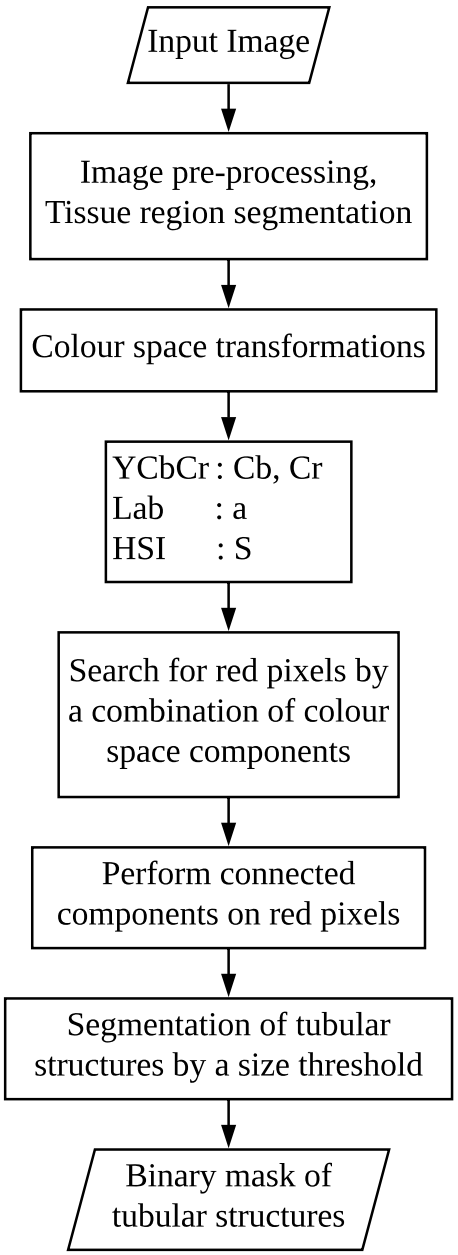


Figure 3.3 The steps taken in the derivation of the binary mask for tubular structures in the biopsy images.

the effect of including this component as a negating factor on the redness of a pixel is also investigated. The saturation (S) component of the *HSI* colour space indicates the 'colourfulness' of the *RGB* image, therefore it is included to distinguish between red and white pixels. The resultant red intensity image (*Red*) is calculated as

$$Cb_{invert} = 1 - Cb_{intensity} \tag{3.12}$$

$$Cb_{thresh} = \begin{cases} 1 & for\, Cb_{invert} > Otsu\, threshold \\ Cb_{invert} & otherwise \end{cases} \tag{3.13}$$

$$Red = Cr_{intensity} \times a*_{intensity} \times saturation \times Cb_{thresh} \tag{3.14}$$

It was observed that the hue variations in the redness of the stained biopsy image are still present despite intensity normalisation. This results in variations in the average redness across all images. Therefore, an Otsu threshold is employed to determine the threshold value for *Red* in order to produce a binary mask of red pixel for the biopsy image. Figure 3.4 illustrates the steps taken to produce the mask of red pixels.

An 8-connected components analysis is performed on the *Red* pixels image to acquire size and shape information of the groups of red pixels in the image. In other words, identified red pixels are grouped together if they are within the neighbourhood of each other. The 8-neighbourhood is defined as $N_8(p) = \{(x+1,y),(x-1,y),(x,y+1),(x,y-1),(x+1,y+1),(x-1,y-1),(x+1,y-1),(x-1,y+1)\}$. The size of each of these connected groups of pixels are then obtained from the connected component analysis for the extraction of tubular structures. Tubular structures are defined as connected red structures that are larger than 1000 pixels. This size threshold is the minimum size of a tubular structure, as empirically determined by pathologists in this study. The segmented regions are morphologically filled to extract white lumen spaces together with

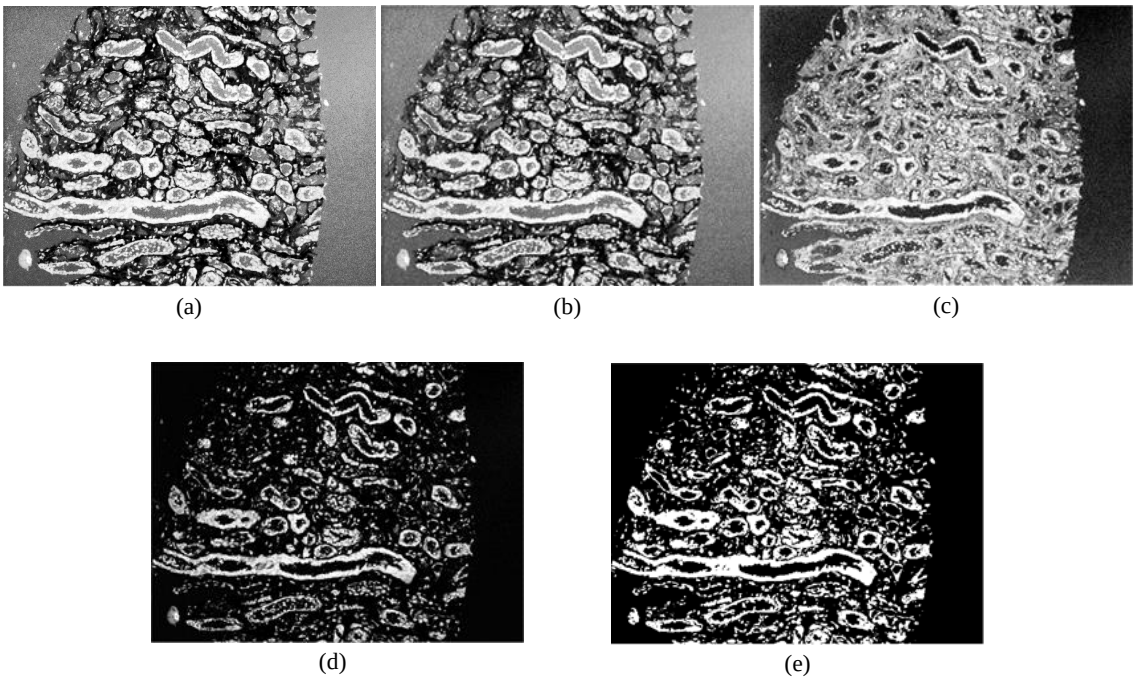


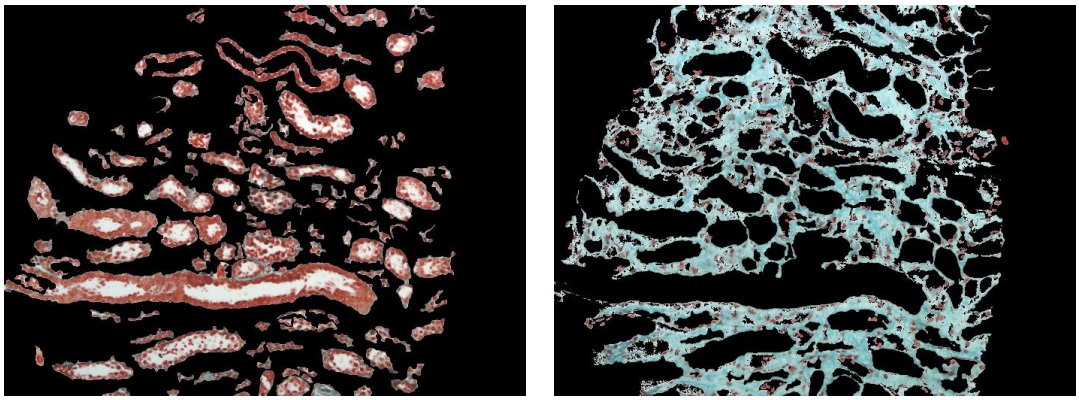
Figure 3.4 The *Red* pixels identification process. (a) *Cr* intensity image (b) *a** intensity image (c) Colour *saturation* image (d) The resultant red intensity image (*Red*) as a combination of *Cr*, *a**, *saturation*, and *Cb* threshold images (e) *Red* pixel mask obtained by the Otsu thresholding.

the tubular walls. These tubular structures are subsequently extracted by identifying the pixels corresponding to these structures.

In biopsy images with only tubular structures present, the IF structures are deduced to be the leftover region in the biopsy tissue area after tubular structure extraction. An example of the tubular structures extracted from the biopsy image is shown in Figure 3.5. The experiment assessing the accuracy of the tubular segmentation against a ground truth image set is documented in Chapter 3.3.4.

3.2.2 Lumen and cellular spaces segmentation

The lumen of a renal tubule is the interior space of the tubule through which the urine passes through. It is a part of the tubular structure and presents itself as an empty



(a) (b)

Figure 3.5 An example of (a) extracted renal tubular structures, and (b) the fibrotic region of biopsy sample after the removal of tubular structures.

space in the biopsy sample. Ideally, the lumen would be extracted along with the tubular structures, except in cases where the tubular walls are fragmented. Similarly, the cellular spaces found scattered in the interstitium matrix are the remnants of cell cytoplasm which also show up as empty unstained spaces in the biopsy images. These cells are not part of the IF and thus would not be included in its quantification. Therefore, the lumen and cellular spaces are simply grouped and identified as white spaces in the images.

Blank white spaces in the image are determined by selecting the highest intensity pixels from the k-means biopsy image, with k=3 reflecting the three main colours (white, blue-green, and red) present in the image. To filter out possible light blue pixels wrongly classified by the k-means image, the validity of these pixels is confirmed by examining their normalised RGB values. Only pixels with all three RGB values within 5% range of each other are accepted as white pixels, as light blue pixels will have a significantly lower R value than white pixels.

3.3 Experimental validation

3.3.1 Biopsy sample images

The kidney biopsy image dataset used in this research has been obtained from the Sultanah Aminah Hospital, Johor Bahru. Biopsy slide records from a total of 70 patients with a range of renal disease and disease severity were acquired from the hospital. These biopsy slides were observed under the Olympus BX53 system microscope with the objective magnification of 10×. They were then captured section-by-section by the Olympus DP21 camera for microscopes with a pixel resolution of 2448×1920 in JPEG format, amounting to 740 colour images to serve as the base dataset for the purpose of this research.

The kidney biopsies acquired have been chemically stained by the MT, a common chemical stain utilised in the biopsy slide preparation procedure to allow the direct observation of certain structures in the tissue sample. This chemical stain reacts with the collagen in the tissue to produce a blue-green hue, while the muscular tissues are stained red [7]. This allows the pathologists to distinguish the IF, which is essentially a collagen structure, from the rest of the tissue. A few example images of the biopsy samples are given in Figure 3.6, illustrating the different relative scales of severity in the kidney diseased states. For details on the compilation of the testing image dataset and the ground truth image dataset of manually annotated IF regions, the reader is referred to Appendix A.

3.3.2 Image pre-processing and tissue region segmentation

Non-uniformities in the biopsy sample due to variations in the staining skills, scanning equipment, and presence of folds in the samples are common in practice as illustrated in

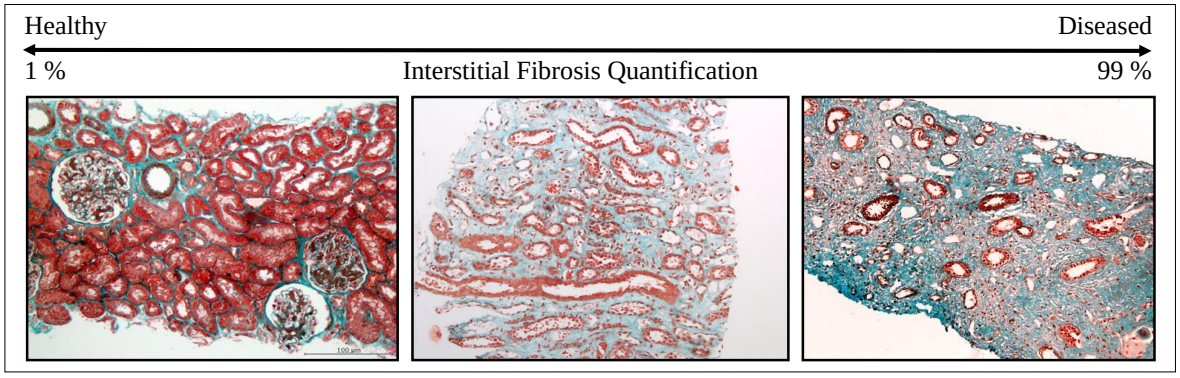


Figure 3.6 Example images demonstrating the stages of renal degradation and corresponding IF amount in renal biopsy samples.The increase in blue-green hue and the corresponding reduction in red muscular tissues indicates degradation.

Figure 3.7. These variations can cause inconsistencies in the detection and segmentation of the fibrosis regions across different biopsy images. While it is nearly impossible to account for all possible variations that can arise in a biopsy dataset, the contrast is maximised in each image to provide a fair basis for comparison between different images. For the purpose of colour segmentation, the histograms of *RGB* colour channels for each biopsy image were stretched. The range of colour values present in the colour image is expanded linearly to fill the maximum range of the colour histogram. This process maximises the contrast, thus sharpening the colour edges in the image.

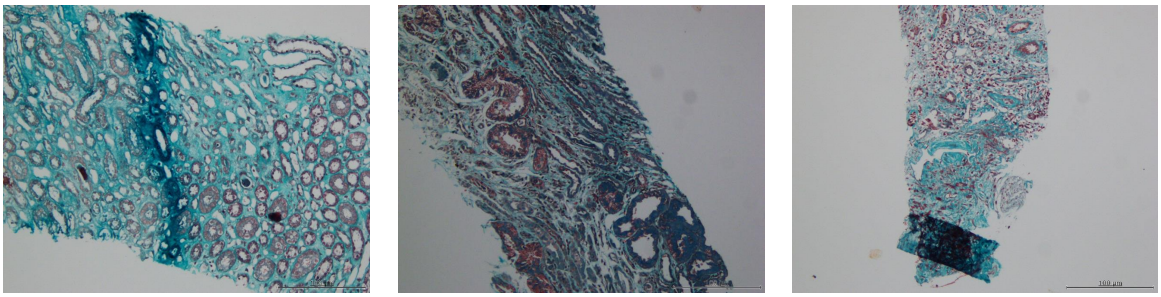


Figure 3.7 Examples of variations and artefacts in the biopsy image dataset. (a) Inconsistent staining (b) Uneven illumination (c) Folds in the biopsy sample.

Next, the white background area has to be segmented in order to compute the total biopsy tissue area for IF quantification (expressed as a percentage of biopsy tissue area). Compared to the blank background region, the tissue region shows significantly greater textural details. This translates to a higher energy area (in signal processing sense) and can thus be differentiated easily from the background region by exploiting the texture information in the image. Unfortunately, due to uneven illumination conditions in the imaging process, the background in some of the biopsy images may present a noisy speckled texture, thus rendering the assumption of a low energy background invalid.

To mitigate this problem, only the colour edges are detected based on colour invariance [301] in the image instead of greyscale intensity edges. The reader is referred to Appendix C for details on the colour invariance method. This method works well because the tissue region has a clear contrast in colour with the white background. For the inspection of biopsy images, an assumption of matte, dull surface, planar object, and an equal energy and uniform illumination is made of the colour images. This colour invariance has a property of an edge detector, where the edges present in the image based on colour gradient are derived independent of intensity level. Morphological operations were performed on the binary image of the colour edges to allow the extraction of the tissue area in the form of a binary image. Figure 2.4 illustrates an example of this operation.

3.3.3 Experimental setup

The accuracy and applicability of the IF quantification by tubular segmentation in kidney biopsy images are tested through three experiments as listed in Table 3.3.

140 scanned images of renal biopsy samples taken from 24 patients at Sultanah Aminah Hospital, Johor Bahru were selected from the dataset specified in Appendix A. For the experiments in this section, the image dataset is limited to include only biopsy

Table 3.3 The list of experiments conducted for validating the performance of the proposed method of segmentation and quantification system.

Experiment	Description	Aim
1	Computerised quantifications of IF based on the established method of colour segmentation (thresholding and clustering) and tubular structures segmentation are performed on 20 renal biopsy images (10 healthy, 10 diseased) with manually segmented ground truth.	To obtain and compare the quantitative accuracy of the two methods (colour segmentation and tubular structures segmentation) of computerised IF quantification system.
2	Computerised quantification is performed on 140 renal biopsy images.	To determine the consistency of the computerised IF quantification system with the pathologists' quantification results.
3	Conventional histopathologic evaluation of IF is independently performed by five skilled pathologists, on the same 140 renal biopsy images used in Experiment 2.	To investigate the inter-observer variation in traditional quantification of IF by visual evaluation.

images without glomeruli and large blood vessels. The reason behind the selection of this subset of biopsy images is because the images taken were sections of the biopsy specimen and not of the entire biopsy slide. Based on observation of the dataset, images of biopsy sections showing only tubular structures make up approximately half of the entire biopsy image dataset. Therefore, the image dataset can be appropriately split for use in this first half of the research project.

From this image dataset, a smaller set consisting of 20 images in which the fibrotic area is manually segmented was prepared as the ground truth set and verified by a skilled pathologist. This process was conducted through the examination of the selected digital biopsy images and the digital segmentation of regions corresponding to IF. These segmented fibrosis areas were saved for use in the subsequent experiments as validation data. This dataset comprises of 10 images each from relatively healthy and diseased biopsy samples respectively. In order to mimic practical conditions in acquiring

biopsy images, the dataset is inclusive of biopsy images presenting variations such as illumination, stain and tissue preparation variations.

3.3.4 Automatic quantification (Experiment 1)

The conventionally straightforward methods of IF quantification by colour thresholding and clustering were investigated and contrasted against the tissue structure segmentation method. The colour thresholding and clustering methods aim to identify only the colour corresponding to the stained IF areas for quantification, while the tissue structure segmentation method is based on a more comprehensive detection approach. All three approaches were employed to produce a quantification of IF in percentage of the biopsy sample images. The image dataset utilised in this experiment was the 20 image set with manually annotated regions of IF.

The quantification results for colour segmentation of IF region in the various colour spaces investigated, and the quantification method based on tubular structures identification are presented in Table 3.4. The box plots showing the distribution of errors for each of the method are also shown in Figure 3.8. The error displayed by the box plot is calculated as the difference in percentage points between the computerised quantification result and the ground truth in terms of fibrosis percentage.

Comparing the performances of the colour segmentation methods, the thresholding method presents a larger variation in its accuracy than the clustering method. The largest deviation is shown by the *YCbCr* colour space, of which error values typically range from 7% to 25%. In contrast, the *Lab* colour space presents the lowest amount of error, averaging at 11.6%. On the other hand, the clustering method recorded almost equal level of errors, averaging at 12% when compared against the manually annotated ground truth by pathologists for all the individual colour spaces. They also exhibit a large deviation, which can be observed from Figure 3.8(b) whereby the errors can vary

from 5% to 17% from image to image. The large error range is almost comparable to the range of a category defined in the Banff classification, where each category of fibrosis severity is of intervals of 25%. In other words, a quantification error approaching 25% would potentially result in a wrong interpretation between a mild, moderate and severe IF diagnosis. The reader is referred to Chapter 1.2 for a more detailed exposition on the Banff classification categories.

Table 3.4 Performance of IF quantification by colour segmentation and tissue structure identification, with their performance quantified by the mean error when compared to the ground truth.

Quantification method	**Colour space**	**Mean error (%)**
Colour thresholding	*RGB*	15.2
	YCbCr	16.3
	Lab	11.6
Colour clustering	*RGB*	12.1
	YCbCr	12.5
	Lab	12.5
Tubular structures identification	*RGB, YCbCr, Lab, HSI*	6.1

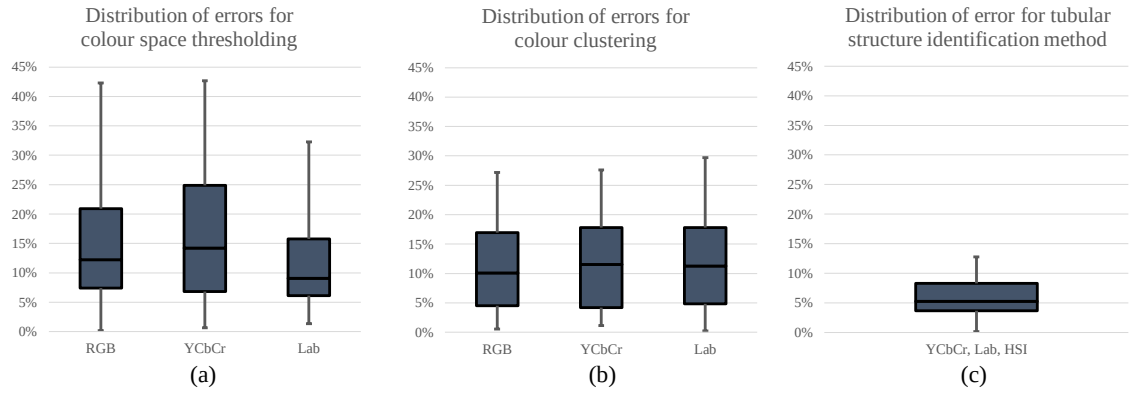


Figure 3.8 Box plots showing the distribution of error for quantification of IF by the (a) colour thresholding method, (b) colour clustering method for the *Lab*, *YCbCr*, and *RGB* colour spaces, and (c) the tubular structure identification method.

On the other hand, the IF quantification by tubular structure identification method produces an improved quantification result compared to the colour clustering method for individual colour spaces. The quantification error is almost half of that obtained by the colour segmentation methods, as demonstrated in Table 3.4. Additionally, this method also shows a considerably lower deviation, as shown in Figure 3.8(c).

The low average error of 6.1% achieved by the tubular structures identification method translates into a lower chance of misdiagnosis in terms of Banff categories, which are of the categorical ranges 0—25%, 26—50%, and >50%. In contrast to the colour thresholding and clustering methods, which have attained higher average errors of 16% and 12%, the tubular structures identification method is much more likely to produce an accurate diagnosis category on the severity of IF according to the Banff classification criterion. Based on observations, the main cause for the higher error rates produced by the colour segmentation methods lies in the fact that the blue-green stained regions are poorly contrasted against the white biopsy background. This have a negative impact on the accuracy of fibrosis region segmentation. However, with the tubular structure segmentation method, the issue is overcome by the detection of colour edges, which helped in a clearer definition of the tissue boundary. Therefore, this step eliminated false negatives and raised the accuracy of the quantification of IF.

This has proven the success of the tubular structure segmentation method by combining the effects of several colour spaces in determining the red pixels in the biopsy images. In other words, the method of IF quantification through the segmentation of all kidney tissue structures (as opposed to only IF) is justified as the direction of this research. A note on the relatively small image dataset with ground truth used in Experiment 1: the size of the image dataset undoubtedly has a certain influence on the accuracy achieved by the tubular structure segmentation algorithm. However, the result has served to prove the superiority of the tubular structure segmentation method by a combination

of colour spaces. This has in turn led to the confirmation of the suitability of the tissue structure identification method in renal IF quantification.

3.3.5 Computerised against visual evaluation quantification (Experiments 2 and 3)

With the intention of presenting a direct comparison between the automated quantification method and the pathologist's common practice of visual estimation, five practising pathologists were invited to participate in a quantification experiment. The proposed automated quantification system was tasked to produce a measurement of the amount of IF in the 140-image dataset described in Appendix A, alongside the five pathologists producing estimates of IF amount on the same dataset. The intraclass correlation coefficient (ICC), Cohen's kappa and a graphical representation of the difference in estimations between the human experts and computer program were the three measures used for calculating the correlation between computer system and pathologists. For an explanation on ICC and Cohen's kappa, the reader is referred to Appendix D.

The statistical analysis of the agreement between the pathologists' scorings and the computer quantification result shows a fairly high amount of consistency, with an ICC of 0.7173. Though, the Cohen's kappa only achieved a value of 0.0145 (p=0.7539). The mean of the percentage ratings given by the pathologists was used in this comparison as a generalised model for the pathologists' quantification. For a clearer picture on the agreeability between the computer software and the pathologists, a graph was plotted as shown in Figure 3.9. It is proven in the graph that the automated quantification result mimics the trend of the pathologists' visual scoring result in a general relative scale, despite the presence of a linear shift. This linear shift explains the low Cohen's kappa value, as the exact quantification values between the proposed system and the

pathologists are not equivalent. Yet, a high ICC value proves that both grading sources show consistency in their own respective relative scales.

A later comparison in Chapter 5 has proven that the automated quantification actually exhibits a stronger correlation to the average pathologist than the agreement between the pathologists themselves. This revelation becomes obvious by looking at the comparison between the ICC values for the computer-pathologists and inter-pathologist agreements. The ICC values of 0.72 (computer-pathologists) against 0.62 (inter-pathologist) have proven the degree of adherence to the average pathologist that the quantification system has achieved. The extent of pathologist variability in visual quantification is investigated in Chapter 5.

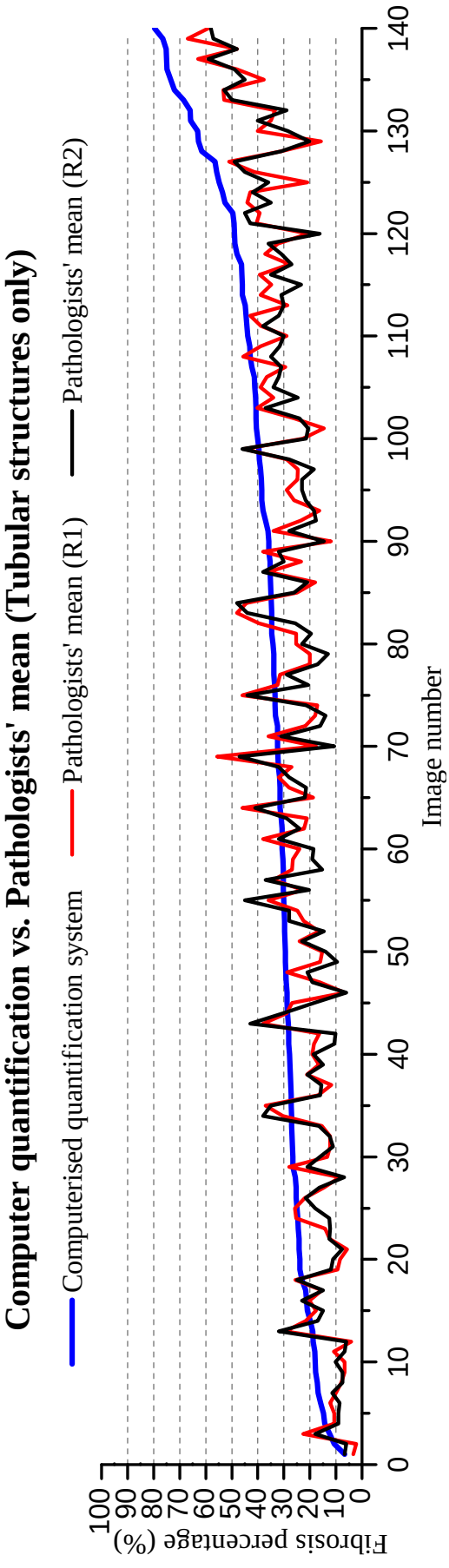


Figure 3.9 Computerised quantification result in comparison with two rounds of the average pathologists' visual quantifications on the 140-biopsy image dataset exhibiting only renal tubular structures.

Chapter 4

Segmentation of Glomerulus and Complete Quantification System Analysis

The IF quantification by kidney tissue structure identification and segmentation has proven to be the more effective method than simple segmentation of IF by colour information in Chapter 3. A direct continuation from that is carried out in this chapter, in which the methods for the segmentation of the remaining structures of interest (Structures 5—8 in Table 1.1) are investigated. These tissue structures are ordered in the frequency of their occurrence in a kidney biopsy tissue, with the least common structure listed last. The proposed algorithm for the quantification of IF by the identification and segmentation of kidney tissue structures is presented in Figure 4.1. The relevant biopsy image dataset containing these structures and their corresponding ground truth images were prepared for performance validation experiments selected from the digital biopsy image dataset. Details of the dataset compiled for testing is recorded in Appendix A.

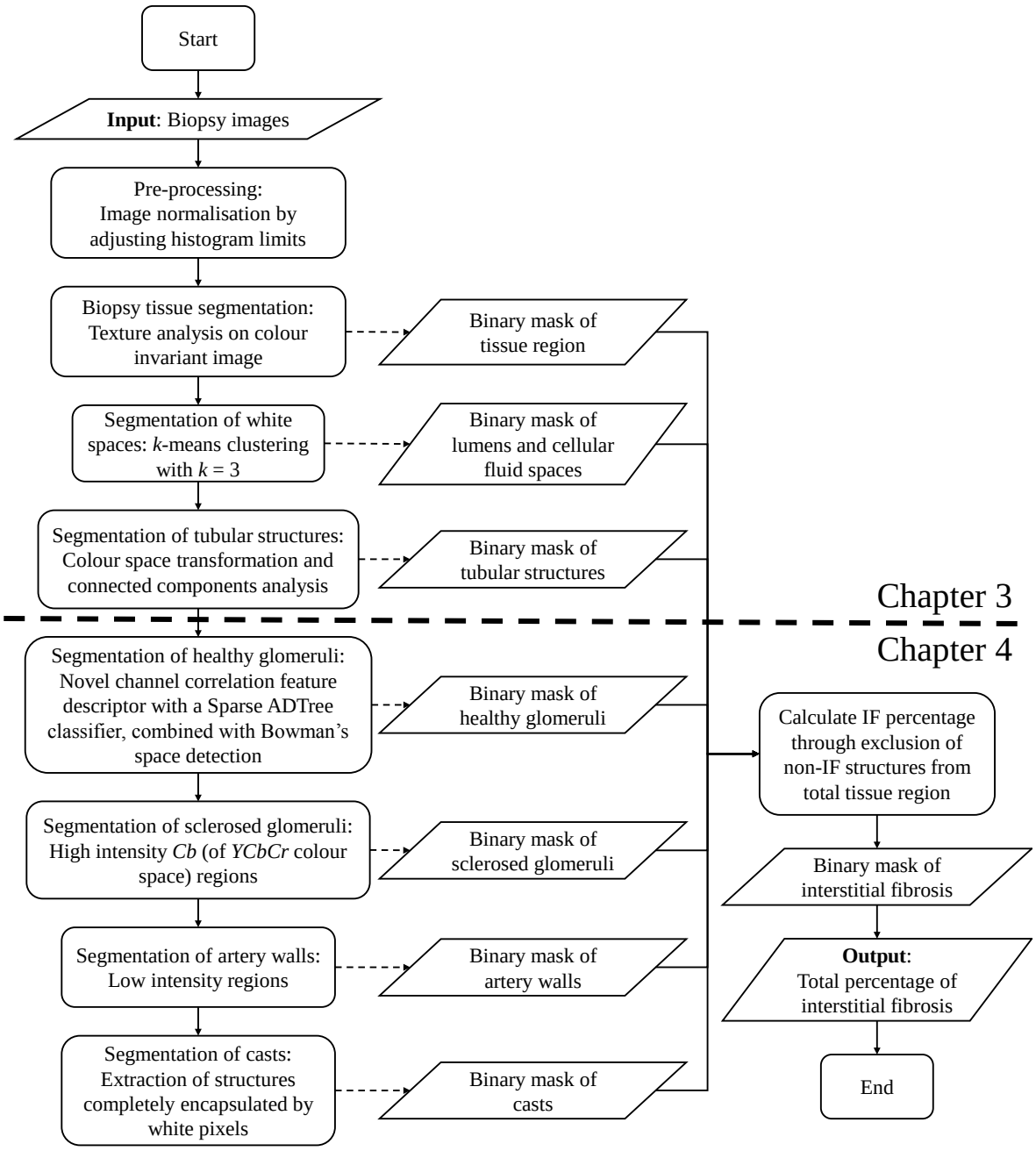


Figure 4.1 Flowchart describing the techniques used in the algorithm for the IF quantification system along with the inputs and outputs specified.

4.1 Healthy glomerulus structure segmentation

The glomeruli are spherical tufts of capillaries in the kidney, and usually appear as circular or oval structures in kidney biopsies. However, their appearance is highly

dependent on the biopsy slicing orientation and preparation, as well as its diseased state. This structure is enclosed by a Bowman's capsule, which leaves a blank space surrounding the glomerulus as the fluid is drained in biopsy slide preparation. Healthy glomerulus structures stained by the MT typically appear darker and have a lower pixel intensity level than other structures in the biopsy slide due to the presence of intraglomerular mesangial cells. Additionally, these intraglomerular mesangial cells present an image texture that is distinctively different from its surrounding tissue structures. Therefore, the use of these characteristics in locating regions of glomeruli in the biopsy image is investigated in this chapter.

Three glomerulus segmentation methods are explored in the following subchapters. Chapter 4.2 details the segmentation method grounded in the search for Bowman's space, while a detection technique based on texture classification of the intraglomerular region is put forth in Chapter 4.3. Lastly, a novel textural classification feature is proposed in Chapter 4.4 as the technique used to generalise the glomerulus appearance for segmentation. An example of the biopsy image showing examples of these structures is presented in Figure 4.2. The large variation in appearance of the glomerulus structures seen in renal biopsy tissues is illustrated in Figure 4.3 in order to provide context for the difficulty in the digital segmentation of this structure.

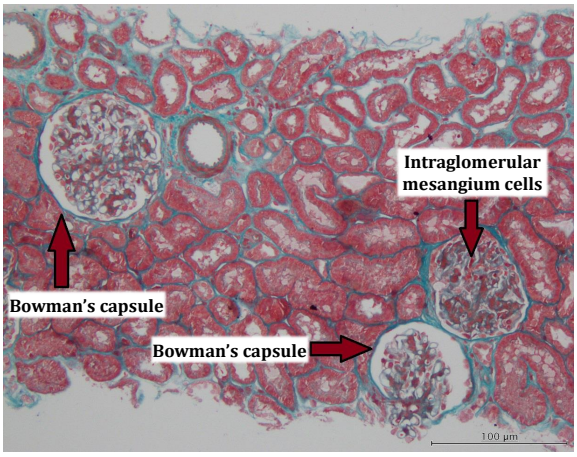


Figure 4.2 An example biopsy image detailing the prominent glomerulus structural features, including the Bowman's capsule and intraglomerular mesangium cells.

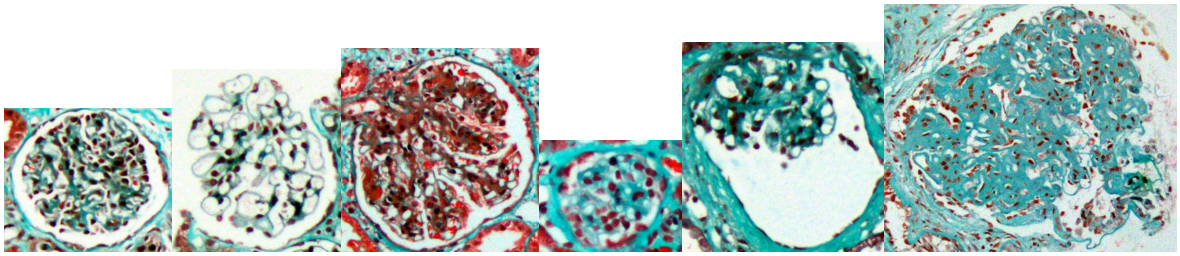


Figure 4.3 The different variations in the appearance of the glomerulus structure in biopsy sample images, highlighting the challenge in segmentation task of the structure. (a-c) Typical glomeruli (d) Glomerulus with no obvious Bowman's space (e) Glomerulus with displaced intraglomerular region (f) Diseased glomerulus.

4.2 Bowman's space search

The segmentation of glomerulus by the identification of Bowman's space is conditional on the identification of candidate glomerulus regions in the biopsy tissue. The shortlisting of probable glomerulus regions is done through the extraction of low intensity areas indicative of intraglomerular mesangial cells in the biopsy tissue. However, low intensity areas of the glomeruli are a relative feature compared to surrounding tissue regions. Therefore, a dynamic threshold for the image intensity is selected through an iterative process with a selection criterion based on the number of low intensity blobs obtained at each level of intensity.

Firstly, the image intensity threshold is incremented in steps from 10% to 50% of the maximum image intensity. At each of this intensity threshold level, the pixels with an intensity value between zero and the threshold intensity are labelled as 1, while the rest are labelled as 0. Next, these pixels with the label '1' are grouped into blobs through a connected component analysis and the number of blobs at each intensity level is recorded. The appropriate 'low intensity' threshold level is then selected as that with the most number of blobs. At this intensity level, morphological operations are performed to locate concentrated regions of the low intensity blobs as probable regions containing glomeruli. The reasoning behind this step is based on the observation that

the glomerulus structure has a typical appearance of a cluster of low intensity blobs. In order to further refine the possible areas of glomeruli, ellipses are fitted onto these low intensity regions, based on the assumption that glomeruli are elliptical in shape. To account for the likely occurrence of close clusters of multiple glomeruli, multiple non-overlapping ellipses are allowed to be fitted onto each low intensity blob. The parameters for the ellipse fitting at the magnification level of 10× are a minimum major axis length of 200 pixels and aspect ratio of 0.8. The potential regions of healthy glomeruli are chosen as the low intensity regions which were successfully fitted with an ellipse.

Normal healthy glomeruli are typically enclosed by the Bowman's capsule, which presents itself as a white space surrounding the glomerular tissues. The appearance of this Bowman's space in the biopsy tissue is naturally subjected to variation in slide preparation. Its detection is executed by dilating the area around the potential regions of glomerulus and mapping it to the white lumen spaces identified in Chapter 3.2.2. The resultant white spaces surrounding these potential glomerulus regions are then attempted to be fitted with an ellipse, as the Bowman's spaces are characteristically elliptical in shape. Identical parameters for the ellipse fitting used in the search for potential glomeruli low intensity regions are applied in this instance. Only successfully fitted Bowman's spaces along with the interior glomerular tissue region are segmented and extracted as glomeruli structure.

4.3 Intraglomerular texture classification

The second method investigated for glomerulus segmentation which is dependent on the shortlisted low intensity regions in the biopsy sample is the textural classification of the intraglomerular region. As opposed to the Bowman's space search method, this

method tackles the segmentation problem in a more robust manner as it is independent on the detection of Bowman's capsule.

Procedures in biopsy sample preparation inevitably introduce variations in the tissue structures, especially the glomeruli. This effect is particularly obvious in the appearance of the Bowman's capsule in biopsy tissues, where its shape can be irregular and contorted. This in turn causes difficulties in recognising its shape by ellipse fitting. However, the interior glomerular tissue remains intact within the confines of the capsule, and has a distinctive texture compared to other parts of the biopsy sample. Therefore, it is proposed to train an rLADTree classifier [243] with textural features in various colour space components to categorise glomerular tissue from the rest of the biopsy. The rLADTree classifier is chosen over other commonly used classifiers (eg. SVM [302]) due to the inherent ability of the rLADTree to perform a selection of the most discriminatory features implicitly. The 606-element textural feature vector used for the rLADTree classifier include *Segmentation-based Fractal Texture Analysis* (SFTA) [303], GLCM [185], *grey-level run length matrix* (GLRL) [304], histogram variance of colour components, and size of the connected components of white pixels. A dataset of more than 500 glomeruli images was manually segmented to serve as the training set of the rLADTree (refer to Appendix A).

With this classifier trained, it is then used to identify the potential glomerulus regions which exhibit textural properties of the typical glomerulus. This is carried out by feeding the calculated textural features of each potential glomerular region in the image as input to the rLADTree classifier. Based on the classifier output, confirmed glomeruli structures are then extracted and saved for subsequent processing.

4.4 Feature by correlation between image representations

Contrasting to the previous two methods which are reliant on the detection of low intensity tissue regions, a novel method for generating image classifying features is proposed in this section for the detection of glomerulus, which will be referred to as the channel correlation features from this point on in the document. The feature generated is based on the amount of correlation between the different representations or interpretations of the original image to be classified. This inspection on the correlation between image representations allows the creation of a feature that provides invariance to scale, rotation and translation in an image. It is not restricted to the measure of local homogeneity but also maps the spatial characteristics of the image pixels.

The images to be classified are resized to 300×300 pixels to suit the average size of the glomerulus. This standard size also facilitates the generation of feature vector in subsequent processing stages.

4.4.1 Image representations

The images are transformed into various representations commonly found in computer vision applications. These different interpretations of the original image add dimensions of information extractable from the image for analysis. The types of image representations explored in this paper are:

1. Greyscale intensity

   The most common type of image feature used in computer vision for various object and scene recognition-type purposes. The colour components in the image are converted to intensity values between the range of 0 and 255.

2. Blue colour intensity

Colour information is a useful feature for the processing of digital colour images. This information is especially advantageous in chemically stained biological images, where the stains highlight the structures of interest. For this application of glomerulus structure classification, the blue colour is of one of the main colours of the MT-stained tissue samples. Hence, the intensity of blue colour presents a potentially discriminating feature for classification. The blue intensity is derived from multiple colour spaces to maximise the contrast with other colours in the background. Commonly used colour spaces such as *Lab*, *YCbCr*, and *HSI* are combined as given in Equations (4.1—4.3), where *Cb* and *Cr* are from the *YCbCr* colour space, a^* and b^* are from the *Lab* colour space while the S (saturation) component is from the *HSI* colour space. The white pixels in the image are identified by a k-means algorithm with k=3.

$$Blue_{binarise} = 1 - (histeq(Cr) \times histeq(a^*) \times histeq(S)) \tag{4.1}$$

$$Blue_{binarise} = threshold(Blue_{binarise}) \tag{4.2}$$

$$Blue = Cb \times b^* \times \sim white \times S \times Blue_{binarise} \tag{4.3}$$

3. Size of white connected components

Connected-component labelling is a method for grouping pixels into structures for recognition and segmentation. It is useful in inferring the spatial properties of regions with similar intensities, which are conventionally called 'blobs'. In this paper, the glomerulus structure images share a common feature of having small, white, hole-like structures scattered relatively evenly across the glomerulus. Hence, the sizes and distribution pattern of these white spaces in the image is a suitable image representation for the dataset investigated.

The white pixels are firstly identified by a k-means operation on the original colour image, with k value of 3 representing the three main colours in the biopsy images. Subsequently, the connected components of the white pixels are obtained along with the size information for each of them. The image representation is then created as the connected component image with each component labelled with the white structure sizes in terms of number of pixels.

4. Wavelet decompositions (1st to 4th orders)

   The wavelet transform is a popular multiresolution image representation method typically used for texture analysis. It offers a scale-invariant representation of an image by decomposing the image to lower resolutions, thus obtaining the corresponding approximate images [189]. Texture analysis employing wavelet decompositions is commonly carried out by examining the first and second order statistical characteristics of the approximate images. For example, the mean, variance, skewness, and kurtosis can be calculated from the approximated images, while the contrast, correlation, homogeneity, and entropy can be obtained from the co-occurrence matrix derived based on the approximate images [305].

   In this paper, the Haar wavelet decomposition of the blue colour intensity image was examined [306]. Four levels of decompositions were carried out, with the subsequent lower resolution images resized to the original image size. The horizontal, vertical, and diagonal detail coefficients of the wavelet decomposition were obtained. These detail coefficients were then combined in a root of the squares method to produce the wavelet decomposition representation image for that level of decomposition.

5. Gradient magnitudes of wavelet decompositions (1st to 4th orders)

   Finding gradient magnitudes from an image essentially highlights the edges and boundaries in the image. To extract information on the shapes of structures

present in the images, the gradient magnitude is calculated from the wavelet decomposition images. This enables the discovery of object boundaries at different resolutions, thus providing an indication to the complexity and randomness in the image as it is progressively blurred. The Sobel gradient operator was used in this study.

6. Intensity of colour invariance image

   Object edges in images are usually obtained based on the greyscale intensity of the image. However, chemically-stained biological samples require a colour image representation to provide useful information in the demarcation between different structures. To accurately discriminate between colours of distinctly different states, Geusebroek et al. [301] derived a set of invariant expressions for *RGB* images that takes into account the varying imaging conditions in practical cases. Effectively, this property acts as an edge detector specifically for colour images. The derivation of the colour invariance image is provided in Appendix C.

   As a basis of comparison, the values for all these image representations are normalized to a range falling between 0 and 1.

4.4.2 Channel correlation feature extraction

The typical feature images used for image segmentation and object recognition as described in the previous section (Chapter 4.4.1) each presents a specific type of information. Each image representation is a version of the original image after undergoing a transformation operation. Therefore, there exists a correlation between the different image representations which can be quantified and exploited for classification.

Aside from using the Pearson and Kendall correlation coefficients as a measure of correlation between the image representations, several descriptive statistics for univariate

analysis are also investigated and compared between two different image representations. The rationale for this being that the correlation coefficients itself are a derivation from a combination of univariate statistical features. For example, the covariance, variance, and standard deviation of an image measures the various properties of the distribution of its pixel values. When the product of these measures between two different images are obtained, it quantifies the degree of dependency between the two distributions. Therefore, the product of univariate statistical values is computed as a method for quantifying the relationship between two image representations in this study. A table summarising the operations employed for correlating the different image representations is given in Table 4.1. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ represent any two image representations derived in Chapter 4.4.1.

Table 4.1 The definitions of statistical features used to investigate the correlation between image representations, with $\boldsymbol{A}$ and $\boldsymbol{B}$ representing any two image representations derived in Chapter 4.4.1.

Statistical feature	**Description**
Pearson correlation	Measures the linear correlation between two image representations. $$\rho(\boldsymbol{A},\boldsymbol{B}) = \frac{cov(\boldsymbol{A},\boldsymbol{B})}{\sigma_{\boldsymbol{A}}\sigma_{\boldsymbol{B}}} \quad (4.4)$$ where $\sigma_{\boldsymbol{A}}$ and $\sigma_{\boldsymbol{B}}$ are the standard deviations of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively.
Kendall correlation [307]	Measures the rank correlation between two image representations. $$\tau = \frac{n_c - n_d}{N(N-1)/2} \quad (4.5)$$

where n_c and n_d represent the number of concordant and discordant pixel value pairs respectively, while N is the total number of pixels in the image.

Covariance

$$cov(\boldsymbol{A}, \boldsymbol{B}) = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{A}_i - \mu_{\boldsymbol{A}}) \times (\boldsymbol{B}_i - \mu_{\boldsymbol{B}}) \tag{4.6}$$

where $\mu_{\boldsymbol{A}}$ and $\mu_{\boldsymbol{B}}$ are the means of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively.

Variance

$$V_{\boldsymbol{A}} = \frac{1}{N-1} \sum_{i=1}^{N} |\boldsymbol{A}_i - \mu_{\boldsymbol{A}}|^2 \tag{4.7}$$

where $\mu_{\boldsymbol{A}}$ is the mean of $\boldsymbol{A}$.

The variances of two image representations are correlated through a multiplication operation.

$$V_{Cor} = V_{\boldsymbol{A}} \times V_{\boldsymbol{B}} \tag{4.8}$$

Standard deviation

$$S_{\boldsymbol{A}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} |\boldsymbol{A}_i - \mu_{\boldsymbol{A}}|^2} \tag{4.9}$$

where $\mu_{\boldsymbol{A}}$ is the mean of $\boldsymbol{A}$.

The standard deviations of two image representations are correlated through a multiplication operation.

$$S_{Cor} = S_{\boldsymbol{A}} \times S_{\boldsymbol{B}} \tag{4.10}$$

Mean

$$\mu_{\boldsymbol{A}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{A}_i \tag{4.11}$$

The means of two image representations are correlated through a multiplication operation.

$$\mu_{Cor} = \mu_{\boldsymbol{A}} \times \mu_{\boldsymbol{B}} \tag{4.12}$$

1st quartile The 1st quartiles of two image representations are correlated through a multiplication operation.

$$Q_{Cor} = Q_{\boldsymbol{A}} \times Q_{\boldsymbol{B}} \tag{4.13}$$

Median The medians of two image representations are correlated through a multiplication operation.

$$M_{Cor} = M_{\boldsymbol{A}} \times M_{\boldsymbol{B}} \tag{4.14}$$

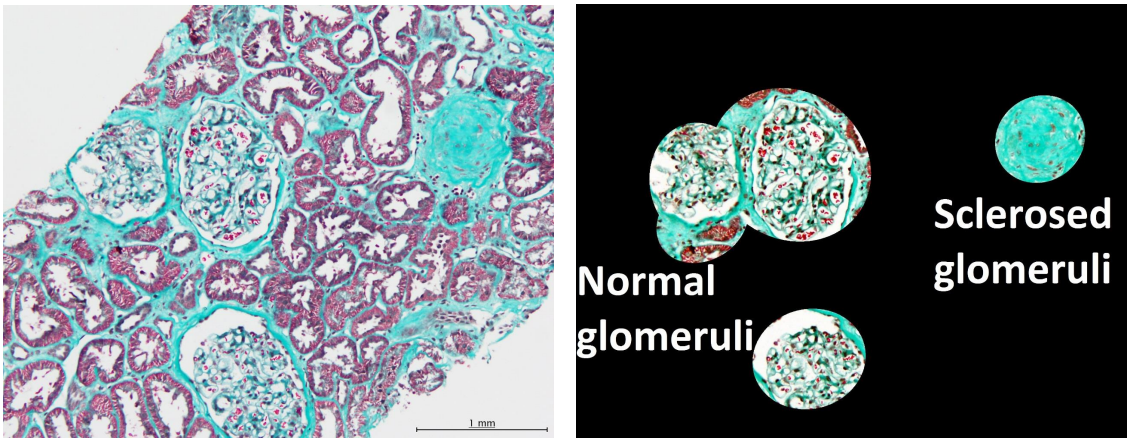
Entropy The entropy of two image representations are correlated through a multiplication operation.

$$E_{Cor} = E_{\boldsymbol{A}} \times E_{\boldsymbol{B}} \tag{4.15}$$

These nine features are calculated for all possible combinations pairs of image representation images described in Chapter 4.4.1. Therefore, there are ${}^{12}C_2 = 66$ pairs of image representations available for the calculation of correlation measures. Since each feature generates a single value, except the covariance being a 4-element feature (from its 2-by-2 matrix), each pair of image representation would produce a 12-element feature vector. These features are then used for subsequent classification by a machine learning algorithm, such as the Sparse ADTree.

4.5 Sclerosed glomerulus structure segmentation

As the kidney function deteriorates, it causes scarring and hardening (sclerosis) of the glomeruli structures in the kidney tissues. The appearance of these sclerosed glomeruli are drastically different compared to the healthy glomeruli. Once stained by the MT, sclerosed glomeruli appear bright blue as they now consist of scarring tissues, similar to that of IF. However, the sclerosed glomeruli tissue is not included in the quantification of IF as this structure is a separate entity. The difference in their appearances is demonstrated in Figure 4.4.



(a) (b)

Figure 4.4 An example of glomeruli segmentation. (a) Sample biopsy image (b) Extracted normal glomeruli and sclerosed glomeruli structures.

The extraction of this structure is carried out by identifying high intensity levels in the Cb component of the $YCbCr$ colour space. This is due to the sclerosed glomeruli's uniformly bright blue appearance, which serves as a useful distinguishing feature for the structure. Two validation steps after the identification of the bright blue pixels are necessary. The first step is the connected component analysis which clusters the identified blue pixels and checks for their sizes. Only connected components with a size larger than 1000 pixels (the minimum size of a structure to be considered as a glomerulus) are kept. Secondly, similar to the healthy glomeruli, the sclerosed glomeruli also appear in a roughly circular shape. An ellipse fitting process is performed onto each of the connected component with the parameters of the minimum major axis length of 400 pixels and aspect ratio of 0.8. Finally, all the pixels in the connected components satisfying these two validation steps are extracted as sclerosed glomeruli structures.

4.6 Artery structure segmentation

The renal artery is made up of smooth muscles which are stained red by the MT, and a hollow lumen at its centre. Due to its tubular nature, the shape it presents in a renal biopsy can vary between circular and a long tubular structure. Their appearance in the biopsy tissue depends on the slicing orientation of the tissue, with their sizes ranging from the size of the tubular lumen to spreading across the entire length of the tissue sample. The irregularity of their shape proves to be a difficulty in segmentation by spatial features. However, a distinguishing characteristic of the artery wall is its smooth muscular tissue structure, which appears as wisps of dark red tissue structures in the biopsy images. Therefore, a detection of low brightness intensity level regions in a roughly tubular form would highlight artery structures. A similar validation step as that performed in the segmentation of sclerosed glomeruli is carried out in this case by connected component analysis. These dark or low intensity regions are checked against

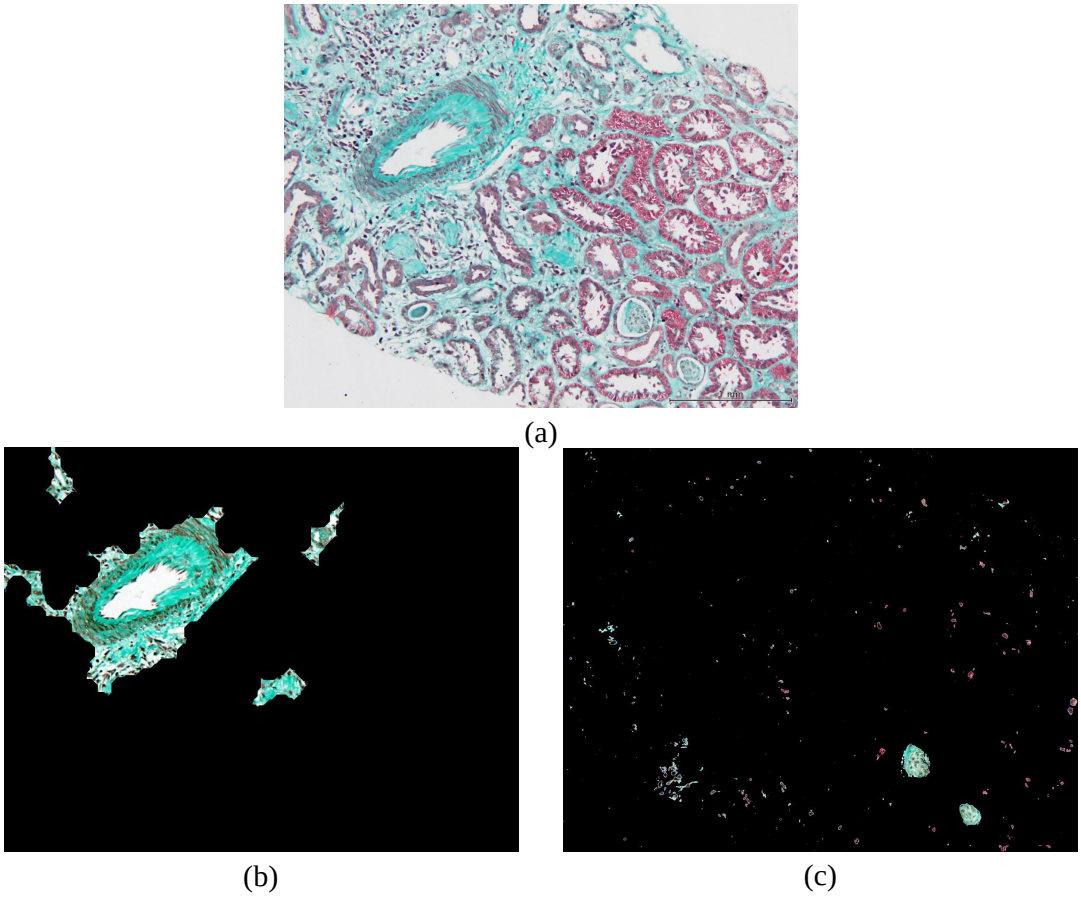


Figure 4.5 Example of arteries and urinary casts segmentation. (a) Sample biopsy image (b) Extracted artery structure (c) Extracted cast structures.

a minimum size threshold of 1000 pixels. This step helps retain actual artery structures while eliminating noise and false targets. A sample image of artery segmentation is given in Figure 4.5(b).

4.7 Urinary cast structure segmentation

In more severe cases of renal function deterioration, excess urea and blood are found in the patient's renal tubules. When their biopsy samples are taken and stained by the MT, this pathological condition shows up in the biopsy sample as urinary casts. Urinary casts can come in the form of protein casts and blood casts. Protein casts appear as blue patches while blood casts are seen as red fragments in the MT-stained

biopsy samples, as shown in Figure 4.5(c). The common feature of these casts is that they are only found on the interior of renal tubules, surrounded by the tubular walls. Since any urinary cast completely enclosed by tubular walls would already be previously extracted together with the renal tubules in connected component analysis, this step would remove any leftover cast in renal tubules with fragmented walls. These structures can be extracted by simply searching for structures completely surrounded by the blank white space in between the renal tubules and urinary casts.

4.8 Complete system quantification and classification

The complete system for the computerised quantification of IF in kidney biopsy images by the identification of all kidney tissue structures is produced by combining the segmentation results of all the tissue structures listed in Table 1.1. The binary mask corresponding to the IF structure is deduced by removing the pixels corresponding to tissue structures 3—8 (Table 1.1) from the binary mask of the tissue region obtained in Chapter 3.3.2. With the pixels of the IF identified, the total area of the structure in the biopsy tissue can be calculated as

$$\text{Interstitial Fibrosis } (\%) = \frac{\text{Number of interstitial fibrosis pixels}}{\text{Number of tissue region pixels}} \tag{4.16}$$

4.9 Experimental setup and results

The accuracy and applicability of the IF quantification by tissue structure identification in kidney biopsy images are tested through three experiments as listed in Table 4.2.

146 scanned images of renal biopsy samples taken from 46 patients at Sultanah Aminah Hospital, Johor Bahru were selected from the dataset specified in Appendix A. Unlike

the experiments conducted in the first stage of the research as specified in Chapter 3.3, the image dataset in this section includes the entire range of biopsy images containing all possible kidney tissue structures.

From this image dataset, a smaller set consisting of 20 images in which the fibrotic area is manually segmented was prepared as the ground truth set and verified by a

Table 4.2 The list of experiments conducted for validating the performance of the proposed automated IF quantification system.

Experiment	Description	Aim
1	Classification of the proposed channel correlation feature descriptor and state-of-the-art features by various classifier types.	To validate the feasibility and accuracy of the novel channel correlation feature descriptor in classification tasks against state-of-the-art methods.
2	The three contrasting methods for glomerulus detection are evaluated against an image dataset of 20 kidney biopsy images (10 healthy, 10 diseased) with manually segmented ground truth.	To obtain and compare the quantitative accuracy of the three methods (Bowman's space search, rLADTree classifier, and channel correlation feature classifier) of glomeruli structure detection.
3	Computerised quantifications of IF by the identification of all tissue structures are performed on 40 renal biopsy images (20 healthy, 20 diseased) with manually segmented ground truth.	To obtain the quantitative accuracy of the computerised IF quantification system.
4	Computerised quantification is performed on 146 renal biopsy images.	To determine the consistency of the computerised IF quantification system with the pathologists' quantification results.
5	Conventional histopathologic evaluation of IF is independently performed by five skilled pathologists, on the same 146 renal biopsy images used in Experiment 4.	To investigate the inter-observer variation in traditional quantification of IF by visual evaluation.

skilled pathologist. This dataset comprises of 10 images each from relatively healthy and diseased biopsy samples respectively. In order to mimic practical conditions in acquiring biopsy images, the dataset is inclusive of biopsy images presenting variations such as illumination, stain and tissue preparation variations.

4.9.1 Performance validation of proposed channel correlation feature descriptor (Experiment 1)

The feasibility of the proposed channel correlation feature descriptor in the detection of glomerulus structures is tested on the glomerulus classification dataset as described in Appendix A. The complete framework for the generation of the channel correlation feature along with the validation stage against established image features are presented in Figure 4.6. The channel correlation feature descriptor is compared against conventional classifying features employed in the examination and classification of biological images.

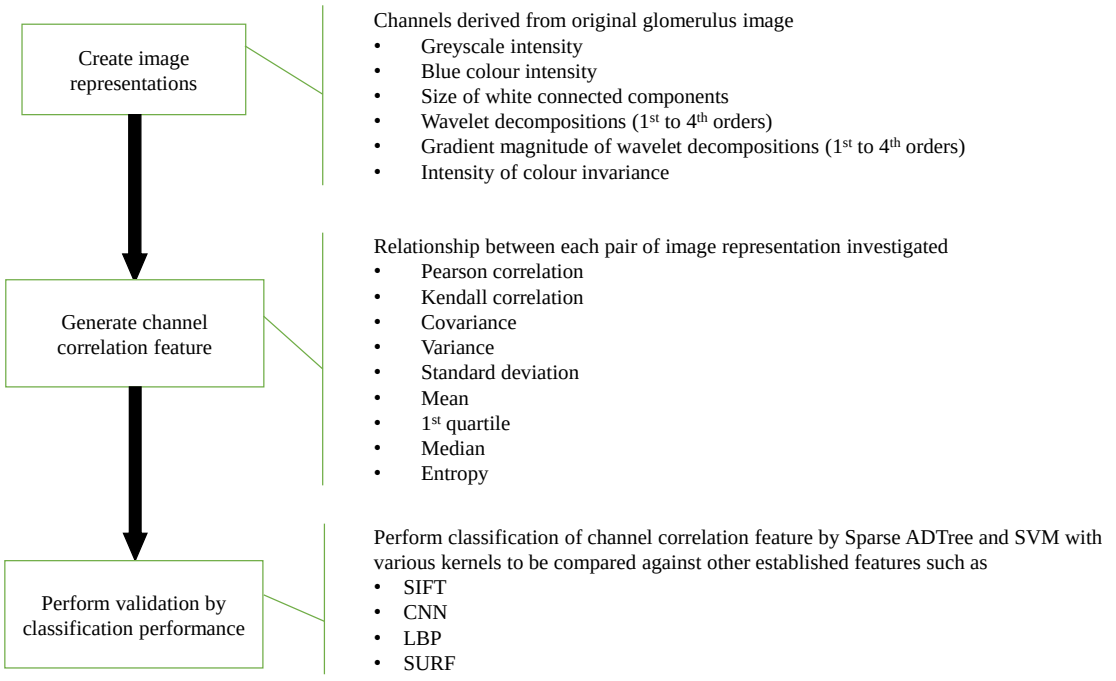


Figure 4.6 The framework for the generation of channel correlation feature, implemented for the recognition of renal glomerulus images.

The SVM and Sparse ADTree employed in this experiment are introduced in Chapter 2.4 (Table 2.1). Since the pattern of the feature data in the feature space is not known at the beginning of the experiment, the most commonly used SVM kernels in image classification are employed in the experiment. These include the simplest linear kernel (of which has been used extensively in text classification) and Gaussian or radial basis function kernel. The conventional features tested include the SIFT, CNN, LBP, and SURF, with the results shown in Table 4.3. Analysis of the classification performance have shown that the proposed feature outperforms the other feature generation methods, with a 100% classification rate with the use of the linear SVM and Sparse ADTree. The performance of the SIFT and CNN feature comes close behind that of the proposed feature, with 99.5% and 98.5% respectively.

SIFT and SURF are local feature extraction methods based on the detection of points of interest, while LBP is based on the local appearance. They are hand-engineered features that extract local details for descriptor generation and classification. CNN, on the other hand, is an unsupervised feature extraction method which is discriminative specific to the given context. In general, the handcrafted features of channel correlation feature and SIFT achieved a higher classification accuracy than the unsupervised feature extraction method of CNN, with slight variations depending on the type of classifier used. SURF, which is a modified faster version of SIFT, is the feature least suited to the use of glomerulus image classification. It is possibly less successful than the SIFT implementation because of its use of integral images as approximations for building an image pyramid, and hence losing a certain robustness to the dataset.

The channel-correlation feature descriptor is suitable for distinguishing between different structures in biological images as it circumvented the randomness in the image by making comparisons between different aspects or descriptions with reference to the spatial position in the original image. Besides, the channel-correlation feature is constructed based on a combination of features. This allows a tunable number of image

Table 4.3 The performance of the proposed feature descriptor against conventional features by various classifiers with 5-fold cross validation, specified in terms of accuracy and the standard deviation in parentheses.

Feature	Sparse ADTree	Linear SVM	Gaussian SVM
Channel correlation	100% (0%)	100% (0%)	81.2% (3.40%)
SIFT	99.4% (0.43%)	99.7% (0.46%)	99.3% (0.58%)
CNN	88.9% (5.88%)	98.4% (0.42%)	98.6% (0.66%)
LBP	89.9% (1.82%)	91.5% (2.21%)	91.2% (2.17%)
SURF	69.6% (4.29%)	77.9% (4.51%)	79.4% (2.36%)

representations to be included by taking into consideration the resource available and complexity required.

As the statistical channel-correlating features are combined into the proposed feature generating framework, its performance has proven superior to that of conventional features, as shown by results recorded in Table 4.3. A note of interest is the comparable performance achieved by SIFT features to the proposed feature. Both the SIFT and proposed channel-correlation feature are scale-invariant, thus hinting at the multiscale nature of the visual cues presented by this image dataset.

4.9.2 Healthy glomeruli segmentation (Experiment 2)

Three contrasting methods for glomerulus identification and extraction are assessed on their segmentation performance on a 20-image dataset with manually annotated glomerulus regions. These are namely the detection by Bowman's space search (Chapter 4.2), the intraglomerulus texture rLADTree classification (Chapter 4.3) and channel correlation feature classification (Chapter 4.4). The result from the glomerulus segmentation experiments involving these methods are recorded in Table 4.4.

Table 4.4 Performance comparison between three glomerular structure segmentation methods on a 20-image ground truth dataset.

Performance index	Bowman's capsule search		Textural rLADTree classification		Channel correlation feature classification	
	Healthy set	Diseased set	Healthy set	Diseased set	Healthy set	Diseased set
Accuracy	55.0 %	54.9 %	75.8 %	62.0 %	71.7 %	74.5 %
Specificity	99.8 %	99.4 %	95.1 %	95.1 %	97.8 %	96.1 %
Sensitivity	10.2 %	10.4 %	56.6 %	28.9 %	45.5 %	52.2 %

As can be seen in Table 4.4, the Bowman's capsule-searching method has a marked lower accuracy than that of the textural rLADTree classification, despite its higher specificity in comparison. This is due to the high rejection rate of potential glomerular regions by the Bowman's capsule searching algorithm, resulting in an almost 100% true negative rate. The flipside to this is the extremely low sensitivity, as correct glomerular structures are rarely surrounded by complete Bowman's capsule space. The presence of this blank space is dependent on the slide preparation process, just as the shape of the glomerulus. Therefore, extraction of glomeruli by the identification of its Bowman's space is a largely hit and miss method. On the other hand, the textural rLADTree classification has achieved a much higher accuracy for both the healthy and diseased set of images. However, the accuracy and sensitivity values are still relatively low, especially that for the more diseased sample set. This can be explained by the increasingly abnormal appearance of the glomeruli in these diseased states, thereby causing an increased difficulty in distinguishing between the glomerulus and its background.

Contrasting between the channel correlation feature and textural rLADTree classification performances, it is apparent that on average, the channel correlation feature classification achieves a higher performance. With the exception of the accuracy and sensitivity on the healthy dataset, the channel correlation feature outperforms the textural rLADTree classification method. The higher success of the channel correlation feature in identifying

glomerulus area can be attributed to the more discriminative information obtained by the relative correlation between different image representations. This adds an extra layer of information on top of textural features extractable from the original image, thus translating into a higher order feature than the basic image textural features utilised by the rLADTree.

Judging by the overall performance, the channel correlation feature classification method appears to be the best choice for glomerulus segmentation. This is at a trade-off for the slightly higher specificity offered by the Bowman's capsule search method, which unfortunately gives an exceedingly low sensitivity. Therefore, the glomerular structure detection method by channel correlation feature classification was selected to be used in the final quantification system.

4.9.3 Computerised quantification of interstitial fibrosis (Experiment 3)

With the incorporation of all the tissue structure segmentation methods into the final complete system for quantifying the IF structure, the proposed system was tested on a 40-image dataset. This image dataset is the combination of both the ground truth datasets used for validating the tissue structure segmentation method (Chapter 3.3.4) and the healthy glomerulus detection method (Chapter 4.3). Thus, the proposed system was assessed over the range of healthy and diseased biopsy samples, with all commonly encountered kidney tissue structures included in the investigation.

The performance of the complete system is shown in Table 4.5, in which the average values for absolute error, accuracy, specificity and sensitivity are recorded. Note that the absolute error refers to the error between actual IF area as a percentage of the total biopsy area and the quantification produced by the automated system. In other words, an absolute error of 10% (percentage points) could indicate that the automated system

quantifies the IF area as 40% of the total biopsy area (as an example) while the actual IF area is only 30% of the biopsy area.

From the table, the mean absolute error is shown to be well below 10%, translating to a reliable quantification for the pathologists in the estimation of IF extent in a biopsy image. This is justified by the fact that a single category of the Banff criteria classification is 25%, as introduced in Chapter 1.2. An accuracy of above 80% for a pixel-by-pixel analysis of the digital biopsy image proves the robustness of the quantification system in a range of kidney deterioration stage that can manifest in the biopsy. A much higher specificity of the system as compared to its sensitivity indicates its much higher capability to correctly identify non-IF structures than IF structure. In other words, pathologists using the system can be reasonably confident in the quantification system's ability to sort through healthy biopsy specimens, while those exhibiting uncertain or uncategorised structures will be flagged for a more detailed analysis from pathologists.

4.9.4 Computerised against visual evaluation quantification (Experiments 4 and 5)

Experiments 4 and 5 in Table 4.2 both involve the same 146-image dataset of biopsy tissues encompassing all tissue structures. The computerised quantification system

Table 4.5 Performance of computerised system against ground truth in 40-image dataset.

Performance index	**Average value**		
	All images	Healthy set	Diseased set
Absolute error (percentage point)	8.8 %	6.1 %	11.5 %
Accuracy	80.1 %	83.5 %	76.6 %
Specificity	88.7 %	89.3 %	88.1 %
Sensitivity	60.5 %	59.0 %	61.9 %

and five pathologists were independently tasked to perform IF quantification on the aforementioned image dataset, which is described in Appendix A. In this section, the degree of agreement between the proposed quantification system and the five pathologists is investigated. The statistical analysis is carried out by employing the ICC and Cohen's kappa (Appendix D), while a graphical representation of the quantification results is provided in Figure 4.7. The comparison between the two rounds of pathologist quantifications for the investigation of intra-observer variation is recorded in Chapter 5.

The ICC value for the computer-pathologists correlation is 0.7043, while the Cohen's kappa value obtained is 0.1295 (p=0.0055). This result echoes the analysis result in Chapter 3.3.5, in which the dataset used was biopsy samples with only tubular structures. An ICC value of 0.70 indicates a sufficiently high amount of agreement between the proposed system and the average pathologist. However, the low kappa value of 0.13 also explains the amount of inconsistency seen in the graph plotted in Figure 4.7. Even though the quantification values of the pathologists and the proposed system do not match up perfectly, the ICC value has demonstrated that they remain in agreement in terms of the relative severity of the IF extent in the biopsy samples. This degree of agreement would be put into the context of inter-pathologist variability in Chapter 5.

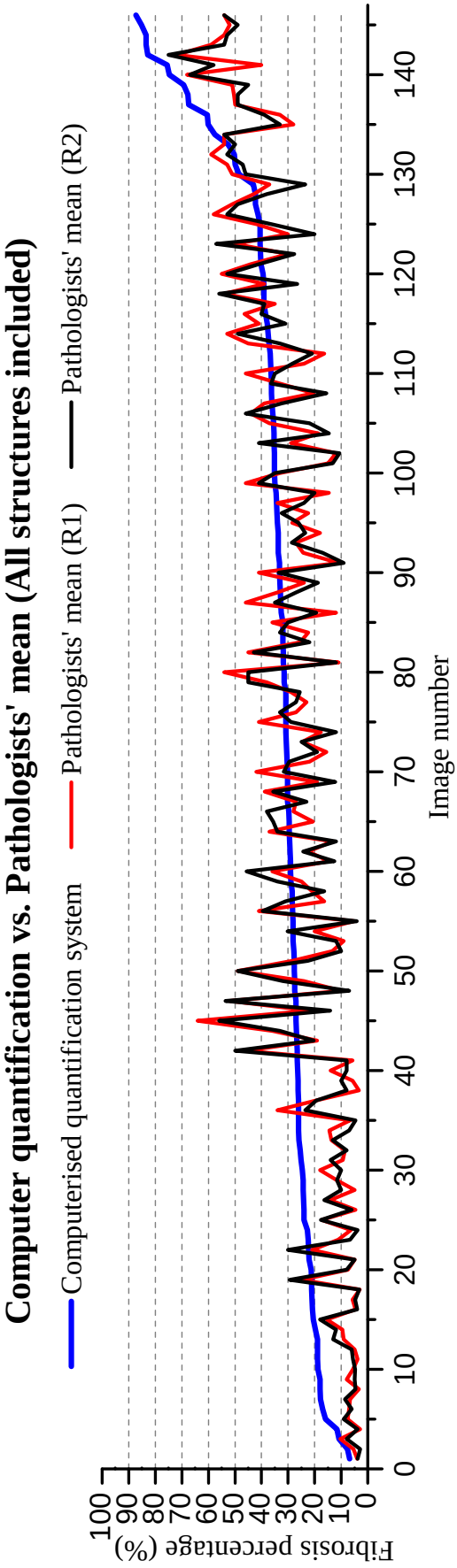


Figure 4.7 Computerised quantification result in comparison with two rounds of the average pathologists' visual quantification on the 146-biopsy image dataset that contains all commonly seen tissue structures.

Chapter 5

Clinical Application

To the best of my knowledge, there has been no existing system that produces an automated quantification of renal IF based on the comprehensive identification and segmentation of all commonly seen kidney tissue structures. The outcome of this research will be a completely new framework for a clinical tool for histopathological laboratories to quantify the extent of renal IF in a patient's biopsy sample. This framework is not only capable of quantification of IF, it also allows the inspection of the rest of the tissue structures as a result of its complete identification approach.

This chapter describes the applicability of the proposed quantification system in the diagnostic process of practising pathologists. The main focus here is the variability in the pathologists' estimations of the IF region in biopsy images and the impact of the proposed quantification system on their variability. Chapter 5.1 details the experimental set up and analysis of the results in which the proposed quantification system was put to use as a diagnostic aide for pathologists. Chapter 5.2 investigates the extent of variability among the pathologists' quantifications. This is followed by the evaluation on the effect the quantification system has on the pathologists' variability in Chapter 5.3. Finally, Chapter 5.4 presents the prototype Graphical User Interface (GUI) for

the proposed IF quantification system, which would facilitate the use of the system for pathologists. The statistical methods employed to quantify the variability between the pathologists are detailed in Appendix D.

5.1 Experimental setup

A total of three experiments were designed to investigate the degree of inconsistency that practising pathologists can exhibit in their diagnoses of renal IF amount in biopsies, and the effect of the usage of the proposed quantification system on this variability. Five pathologists experienced in examining renal biopsy samples and are practising doctors at the Sultanah Aminah Hospital, Johor Bahru have agreed to participate in this study. The experiments conducted are summarised in Table 5.1. Experiments 1 and 2 aim to assess the intra-observer variability in pathologists, using the same image datasets as in Chapters 3.3 and 4.9. This amount of variability would then serve as a comparison to the variability observed when the pathologists are aided by the automated quantification result from the proposed system, which is the result from Experiment 3.

Conventional histopathologic visual evaluation of the kidney biopsy images was independently performed by five skilled pathologists to produce a quantification of IF seen in the images. The pathologists were tasked to quantify the amount of IF seen in a set of biopsy images based solely on their expert opinions. This group of volunteering pathologists are experts with experience in inspecting renal biopsy images ranging from 2 to 24 years. No prior calibration of the pathologists' quantification in the form of standard biopsy images exhibiting a certain severity of IF has been carried out in the experiments. To facilitate a direct comparison between the computerised quantitative result and the pathologists' evaluation result, the quantification was done in terms of percentages of the total biopsy tissue area.

Table 5.1 The list of experiments conducted involving a group of pathologists for investigating pathologist variability and validating the use of the proposed automated IF quantification system.

Experiment	Description	Aim
1	Conventional histopathologic evaluation of IF is independently performed by five skilled pathologists, on the 140 renal biopsy images containing only tubular structures. The experiment is carried out twice with a minimum interval of two weeks between repetitions.	To quantify the intra-observer variation in traditional quantification of IF by visual inspection.
2	Conventional histopathologic evaluation of IF is independently performed by five skilled pathologists, on the 146 renal biopsy images containing all possible renal tissue structures. The experiment is carried out twice with a minimum interval of two weeks between repetitions.	To investigate the intra-observer variation in traditional quantification of IF by visual inspection with the interference from additional renal tissue structures.
3	Histopathologic evaluation of IF aided by automated quantification results by four skilled pathologists on the renal biopsy image set containing all tissue structures.	To measure the effect of having the automated quantification results as a reference on the pathologists' visual quantification results.

Two image datasets were evaluated by the pathologists for the investigation of their variability in quantification of IF, as detailed in Appendix A and Appendix B. These datasets contain images with only tubular structures and all possible renal tissue structures respectively, as was investigated in Chapters 3.3 and 4.9. With the intention of investigating the intra-observer variation, two rounds of evaluations were performed by the same pathologists on the same image dataset over a time interval of at least two weeks. The image dataset evaluated by the pathologists on the second round is the reshuffled image dataset of the one used in the first round of scoring to ensure objectivity in their quantifications. The sufficient time between the two rounds of evaluation ensures

that the pathologists are not influenced by their first round of evaluation for the purpose of investigating their intra-observer variation.

Four performance indices were employed in this chapter to gauge the practicality of the computerised quantification system. These include the mean absolute error, accuracy, specificity and sensitivity. The true positives (TP) are pixels correctly classified as IF, while true negatives (TN) are pixels correctly classified as non-IF structures, and vice versa. To gauge the amount of correlation between the different pathologists and the automatic quantification system, three main statistical methods were employed, namely the ICC [308] and Cohen's (κ_C) and Fleiss' kappa (κ_F) tests. These metrics are the most common statistics used in measuring inter-rater reliability [309] and are described in Appendix D. As a basis for subsequent discussion, the categorical labels for the corresponding ranges of the kappa value introduced by Landis et al. [310] are used to convey the relative strength of agreement between the raters, as illustrated in Figure 5.1.

5.2 Pathologist variability in conventional visual quantification

This section provides an analysis on the agreeability between five independent pathologists in their quantification of renal IF through the routine method – visual evaluation of the biopsy images. Pathologists were tasked to estimate the amount of IF based solely on their expert judgement. The experiments are divided according to the types of structures found in the biopsy tissue samples, namely the ubiquitous renal tubular structures and the rest of the less common renal tissue structures, as described in Chapter 1.1.

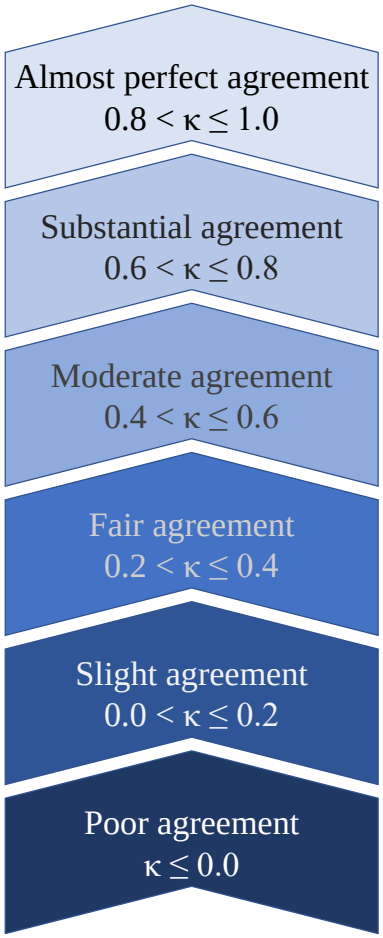


Figure 5.1 The categorical agreement for ranges of kappa values as proposed by Landis et al. [308]

5.2.1 Visual evaluation on biopsy sections with only tubular structures (Experiment 1)

The conventional visual quantification of IF results produced by the pathologists were evaluated on the inter- and intra-observer variability. Results on the evaluation of inter-observer variability are compiled in Table 5.2, while that for intra-observer variability are given in Table 5.3. Assessing the inter-observer variability, the quantification results between different pathologists demonstrated a substantial correlation, with an ICC of 0.6196. However, when assessed by the Fleiss' kappa test with the quantification percentage results rounded to the nearest 10%, the Fleiss' kappa is a drastically low 0.0659 (p=0.0000). This suggests that while the pathologists may agree in different

relative scales of quantification as shown by the ICC, their actual scores for the same images vary significantly enough to not produce a quantification score that conveys the same diagnosis to other pathologists.

Figure 5.2 demonstrates the agreement between each of the five pathologists to the average pathologist (among themselves) on the 140-image dataset with only tubular structures over two rounds in a Bland-Altman plot. In an ideal case of perfect agreement between the pathologists, the Bland-Altman plot would have all points to be well within the two limits of agreement, subject to random variations within two standard deviations about the mean difference. However, as can be observed in Figure 5.2, all five pathologists have evaluation differences that lie well outside the 95% limits of agreement. This indicates a strong disagreement within the pathologists, confirming the Fleiss' kappa test result in Table 5.2.

On the other hand, the Cohen's kappa test is used for evaluating the intra-observer variability in the pathologists, as presented in Table 5.3. The degree of agreement between each pathologist's first and second round of evaluation on the same image dataset has been shown to display only a fair but not strong agreement, as indicated by the relatively low Cohen's kappa value. On average, the five pathologists have recorded a Cohen's kappa value of about 0.27, which corresponds to only a fair agreement based on Figure 5.1. Consistent with the kappa test, the average variation in percentage point of the quantification is shown to go upwards of 10%. To illustrate the point, a pathologist's estimate of IF amount in the same biopsy tissue could change from 20% to 30% of the total biopsy area from one round of evaluation to the next. This result

Table 5.2 The measures for inter-observer variability among the 5 pathologists on the image dataset containing only renal tubular structures by ICC and Fleiss' kappa, where a complete agreement would achieve a value of 1.

Inter-observer variability measure	**Value**
ICC	0.6196
κ_F	0.0659 (p=0.0000)

Table 5.3 Intra-observer variability for 5 pathologists by the Cohen's kappa values where a complete agreement between the raters' first and second scoring would produce κ_C=1, and the mean variation in percentage points.

Pathologist	κ_C	**Mean intra-observer variation (percentage point)**
1	0.3129	6
2	0.2804	7
3	0.2629	7
4	0.2354	10
5	0.2323	9

highlights the need for a more consistent way of quantifying the amount of IF in biopsy images.

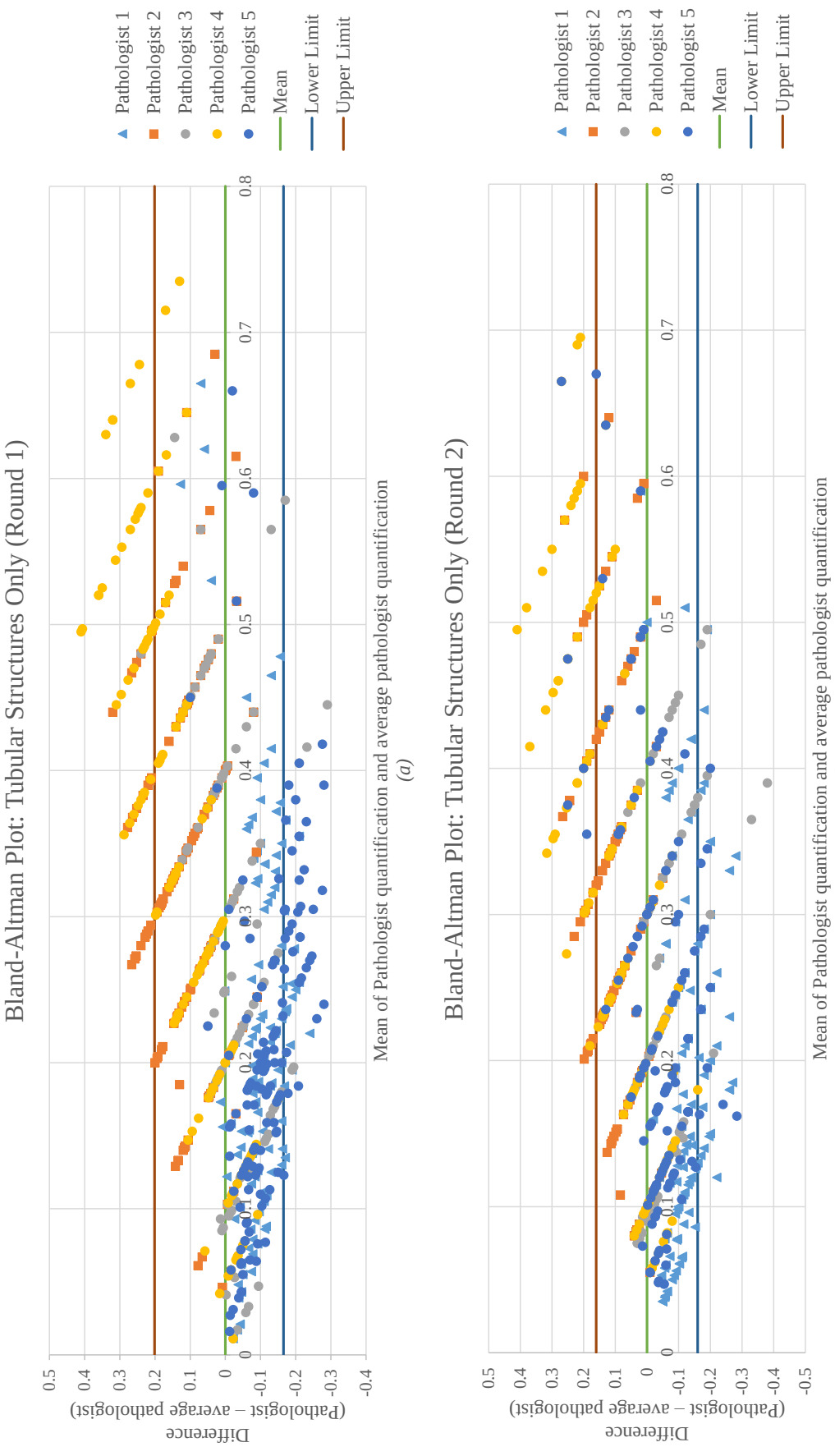


Figure 5.2 The Bland-Altman plots showing the agreement between each pathologist to the average pathologist in (a) the first round, and (b) the second round of pathologists' quantifications on the biopsy image dataset with only tubular structures present.

5.2.2 Visual evaluation on biopsy sections with all tissue structures (Experiment 2)

On the dataset compiled for biopsy images containing all possible tissue structures, the variability in pathologist quantifications exhibits a similar trend as that for the tubular structures-only dataset. In terms of the inter-pathologist variability (results presented in Table 5.4), the ICC for quantification results between different pathologists is 0.7338. This indicates a substantial correlation between the pathologists, while revealing a higher agreement between the pathologists than that in the tubular structures-only dataset. Assessing the pathologists' inter-observer variability by the Fleiss' kappa test, with the quantification percentage results rounded to the nearest 10%, shows an extremely low correlation. The Fleiss' kappa is only 0.1868 (p=0.0000), which is higher compared to that in the tubular-structure-only dataset but still only at a slight agreement level. Once again, this suggests that the pathologists' estimations of the IF area would not produce an agreeable diagnosis between different pathologists and depend heavily on the individual pathologists' own relative scale of judgement.

Figure 5.3 presents the Bland-Altman plots which reveal the degree of agreement between five independent pathologists and the average pathologist (among themselves) on the 146-image dataset with all renal tissue structures. Over two rounds of examination on the same image dataset, it was noted that the pathologists exhibit a similar amount of disagreement as in the previous dataset containing only tubular structures (Figure 5.2). An interesting observation from the graphs is that the pathologists only achieved

Table 5.4 The measures for inter-observer variability among the 5 pathologists on the image dataset containing all possible tissue structures by ICC and Fleiss' kappa, where a complete agreement would achieve a value of 1.

Inter-observer variability measure	Value
ICC	0.7338
κ_F	0.1868 (p=0.0000)

a good agreement between themselves in the cases where the severity of IF is under 10% of the total biopsy area.

The Cohen's kappa values for the five pathologists show a fair to moderate amount of agreement between the individual pathologist's two rounds of quantification, with a highest correlation of 0.47 for Pathologist 3. The average variation in actual percentage values as recorded in Table 5.5 prove that the discrepancy in estimations made on the same dataset in two separate rounds can vary significantly from pathologist to pathologist. Therefore, the proposed automated quantification system is aimed to aid the observers in making a more informed decision and quantification by providing quantifications from an objective standpoint.

Table 5.5 Intra-observer variability for 5 pathologists by the Cohen's kappa values where a complete agreement between the raters' first and second scoring would produce κ_C=1, and the mean variation in percentage points.

Pathologist	κ_C	**Mean intra-observer variation (percentage point)**
1	0.2888	7
2	0.3240	7
3	0.4675	5
4	0.2339	12
5	0.2534	9

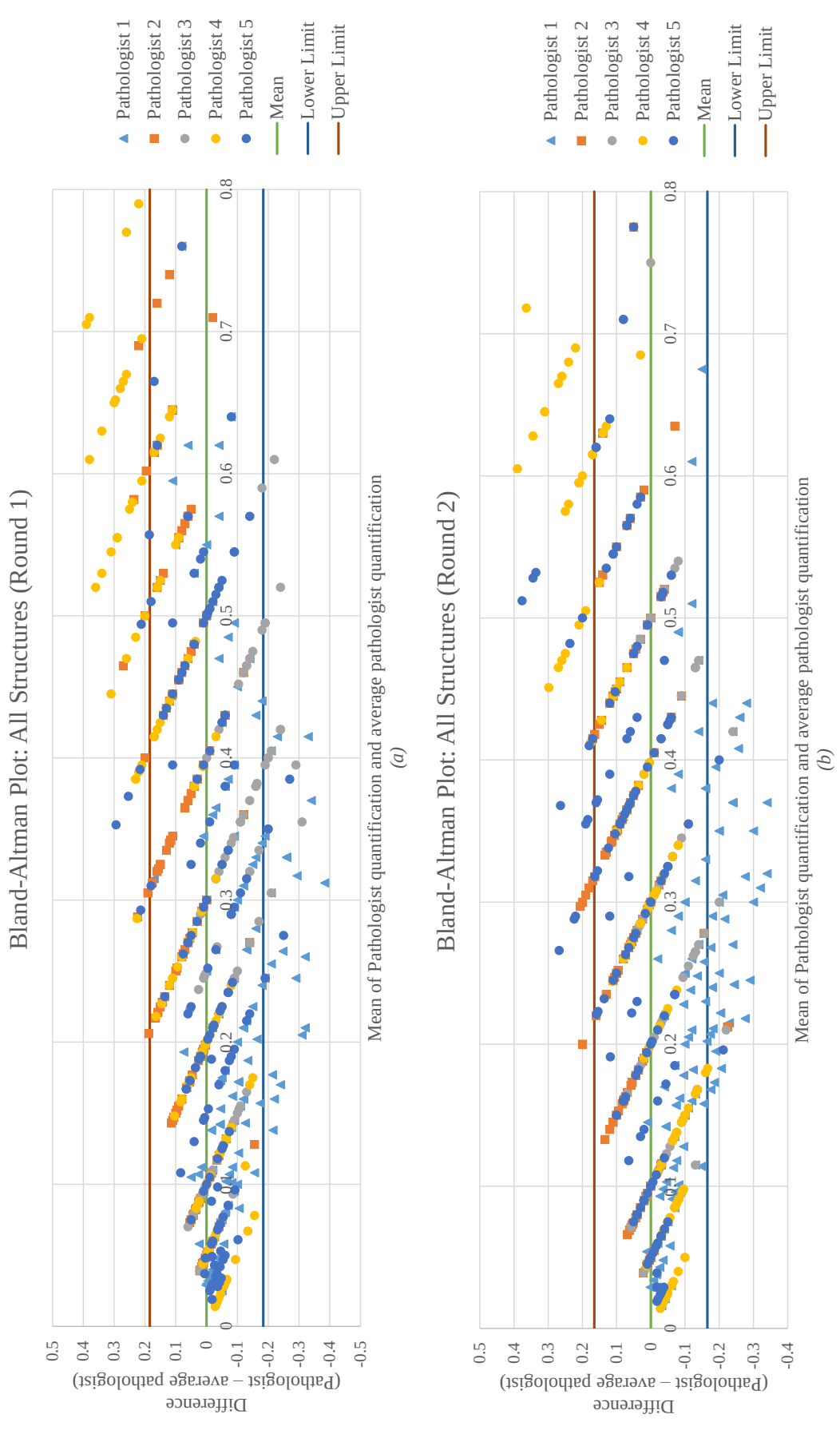


Figure 5.3 The Bland-Altman plots showing the agreement between each pathologist to the average pathologist in (a) the first round, and (b) the second round of pathologists' quantifications on the biopsy image dataset with all possible tissue structures present.

5.3 Automated quantification as a diagnosis aide

This section presents results to the investigation on the variability between four independent pathologists in their quantification of renal IF with the aid of the proposed automated quantification system. In this round of experiments, the pathologists were furnished with the quantification result by the proposed quantification system along with the biopsy images which they were tasked to evaluate. This round of experiment only involves the dataset of biopsy images containing all tissue structures. The results recorded in this section correspond to the experiments listed in Table 5.1. Note that Experiment 3 was conducted with only four pathologists, as one of the pathologists has pulled out from the experiment round. Thus, in this section, any comparison to pathologists' quantification results without using the proposed quantification system is adjusted to only considering the four corresponding pathologists for a fairer assessment.

5.3.1 Computerised against visual evaluation quantification (Experiment 3)

One of the primary purpose of the proposed computerised quantification system is to provide a reasonable basis for discussion between pathologists on the diagnosis of renal IF progression in a patient. In other words, the automated quantification result should be in general agreement with the average pathologist's quantification to pass as a convincing aide to practising pathologists. A previous work by Farris et al. [311] has demonstrated the variability among pathologists from different centres and has suggested the use of an automated assessment as a tool for standardisation. Experiment 3 (Table 5.1) in this study furthers their work by examining the feasibility of having an automated quantification system as a second opinion in clinical diagnosis.

The experiment was carried out in a similar manner as the other experiments involving pathologists. Images of sections of renal biopsy samples were compiled and sent to the group of pathologists for quantification by visual estimation. The differing aspect in this round of experiment (Experiment 3) is the addition of the quantification results produced by the proposed automated quantification system to the image dataset. In other words, the pathologists were supplied with supplementary information of the IF percentage as deduced automatically by the proposed system, along with the image of segmented IF regions in the biopsy image. Having this objective and empirical estimation as a basis for judgement allows the pathologists to make more informed inferences and diagnoses on the biopsy samples. The analysis on the outcome of this experiment is reported in Table 5.6.

Firstly, the results for quantifications without the use of the computer aide by pathologists are analysed. It has been shown that the automated quantification system produces IF estimates that are in a fairly high amount of consistency with the average pathologist, based on the ICC value of 0.704 in Table 5.6. This was also proven in Figure 4.7 (Chapter 4.9.4) where the automated quantification result mimics the general trend of the pathologists' visual quantification result. In fact, this level of consistency is on a comparable level to that among different pathologists, as demonstrated by practically identical ICC values between the two types of correlations investigated (0.704 and 0.695 for the computer-pathologists and inter-pathologist comparisons respectively).

Next, the impact of having the automated quantification as an aide to the pathologists' visual quantification was measured by the ICC and Fleiss' kappa test. Compared to the pathologists' quantification without the aide, having the automated quantification results as a basis for estimation has considerably increased the correlation between independent pathologists. This is demonstrated by the ICC (increase from 0.6950 to 0.7463) and the Fleiss' kappa (increase from 0.1437 to 0.2434) tests carried out. This degree of increased correlation is also reflected in the rise of one agreement category of

Table 5.6 The variability measures between the automated quantification and pathologists' visual quantification and between four different pathologists, with ICC=1 and kappa=1 in the case of perfect agreement.

	Statistical evaluation test	Without aide	With aide
Computer vs pathologists' average	ICC	0.7043	0.8935
	κ_C	0.1295 (p=0.0055)	0.2571 (p=0.0000)
Pathologist vs pathologist	ICC	0.6950	0.7463
	κ_F	0.1437 (p=0.0000)	0.2434 (p=0.0000)

the Fleiss' kappa from a slight to fair agreement, based on the Landis et al. classification in Chapter 5.1.

Finally, it has also been proven that the proposed quantification system displays a correlation to the average pathologist that is higher than the correlation between the pathologists themselves. This is confirmed by the ICC (0.7043 vs. 0.6950 and 0.8935 vs. 0.7463) and the kappa values (0.1295 vs. 0.1437 and 0.2571 vs. 0.2434) in Table 5.6. Furthermore, the automated quantification results have actually exhibited an almost perfect agreement with the average pathologist's evaluation, as the ICC value of 0.89 suggests.

5.3.2 Analysis of quantification in Banff criteria categorisation

In order to provide an assessment on the reliability of the proposed system in a way that is more relatable to the actual diagnostic process of chronic kidney diseases, the quantification results of IF were classified according to the Banff criteria [312]. The three grades of IF as per the Banff classification are:

I. mild (IF $<$ 25% of cortical area),

II. moderate (IF 26—50% of cortical area), and

III. severe (IF >50% of cortical area).

The results from Experiment 3 (Table 5.1) were reassessed in terms of the Banff categories, while the intraclass correlation and Cohen's kappa statistics were calculated to compute the diagnosis consistency between the computer system and the average of the four pathologists. The results of the analysis are compiled in Table 5.7.

Referring to Table 5.7, for experiments involving pathologist quantification by their expert judgement alone, the automated quantification system has actually shown a higher degree of agreement to the average pathologist than the agreement among pathologists themselves. This is proven by the slightly higher computer-pathologists ICC and kappa values than the inter-pathologist ICC and kappa values. This reflects the result presented in Table 5.6 in which the actual quantification values were analysed. In fact, the kappa value for the agreement between the computer and pathologists is higher than that for the inter-pathologist agreement. This fair level of agreement is, in fact, higher than the slight agreement level shown in the case where the absolute values of quantification were compared (Table 5.6).

Table 5.7 The variability measures between the automated quantification and pathologists' visual quantification and between four different pathologists in terms of the Banff categories, with ICC=1 and kappa=1 in the case of perfect agreement.

	Statistical evaluation test	**Without aide**	**With aide**
Computer vs pathologists' average	ICC	0.5904	0.7068
	κ_C	0.3639 (p=0.0001)	0.5538 (p=0.0000)
Pathologist vs pathologist	ICC	0.5913	0.6472
	κ_F	0.3081 (p=0.0000)	0.4345 (p=0.0000)

When compared to the experiment results in which the pathologists were using the proposed computerised quantification system in their evaluation process, the pathologists have demonstrated a significant increase in agreement with other pathologists. This positive effect of the aide on the correlation between the pathologists is shown in Table 5.7, in which a rise in both the ICC (from 0.59 to 0.65) and Fleiss' kappa values (from 0.31 to 0.43) was recorded. Moreover, the ICC value for the computer-pathologists variability has risen up to 0.71 from 0.59 when the pathologists procured the aide in their evaluation. This dramatic spike in agreement between the automated quantification and the pathologists has surpassed that for the inter-pathologist agreement, which has only increased up to 0.65 in ICC value. Furthermore, the Fleiss' kappa value has exhibited a rise in agreement from a fair to moderate agreement category, while the Cohen's kappa value rose from a moderate to substantial agreement. This implies the feasibility of the automated quantification system performing like an independent observer that is as proficient as the average pathologist. The higher correlation between the proposed quantification system and the pathologists also demonstrates the robustness of the computer system to account for the different types of structures in a biopsy image towards IF quantification. This conclusion is justified on the basis of the accuracy of the proposed system when tested against the ground truth dataset in Chapter 4.9.3.

5.4 Graphical user interface clinical tool

As the ultimate aim of this research is to provide a working system for the pathologists as an aide in the quantification of renal IF, the development of a user-friendly GUI adds value to this research. This tool would offer a means for the pathologists or any medical personnel with no background knowledge in the field of computer vision to utilise the quantification system effectively. Additionally, the system also contains functionality for the user to provide feedback on the automated quantification result.

Currently, the prototype of the GUI has been developed in the MATLAB development environment, with the ability to accept digital colour biopsy images and output binary masks corresponding to each of the kidney tissue structures (if present) extracted. The quantification result of the IF structure is also displayed in the GUI as a percentage of the total biopsy area. In order to facilitate feedback from the pathologists, the developed GUI also allows the user to manually correct the binary masks of the tissue structures in the case of an error. The edited images are saved for revisions. The interface for the prototype IF quantification program is shown in Figure 5.4, with instances of operation of the program in Figure 5.5(a—d).

In the future version of the GUI, it is proposed to be developed as a standalone application or a web-based GUI so as to allow remote diagnosis to take place, adapting to the trend of internet medicine. Additionally, the general framework of IF quantification is based on the individual segmentation of all tissue structures in the biopsy image. Therefore, the interpretation of this framework in a working GUI is transferrable to other types of tissue images, provided that the specific tissue structure detection and segmentation methods are derived specific to that tissue type. Since the visual cues extracted from the biopsy images are based mainly on colour, cell structure and image textural information, the adaptation of this GUI onto other tissue types is feasible in a future work.

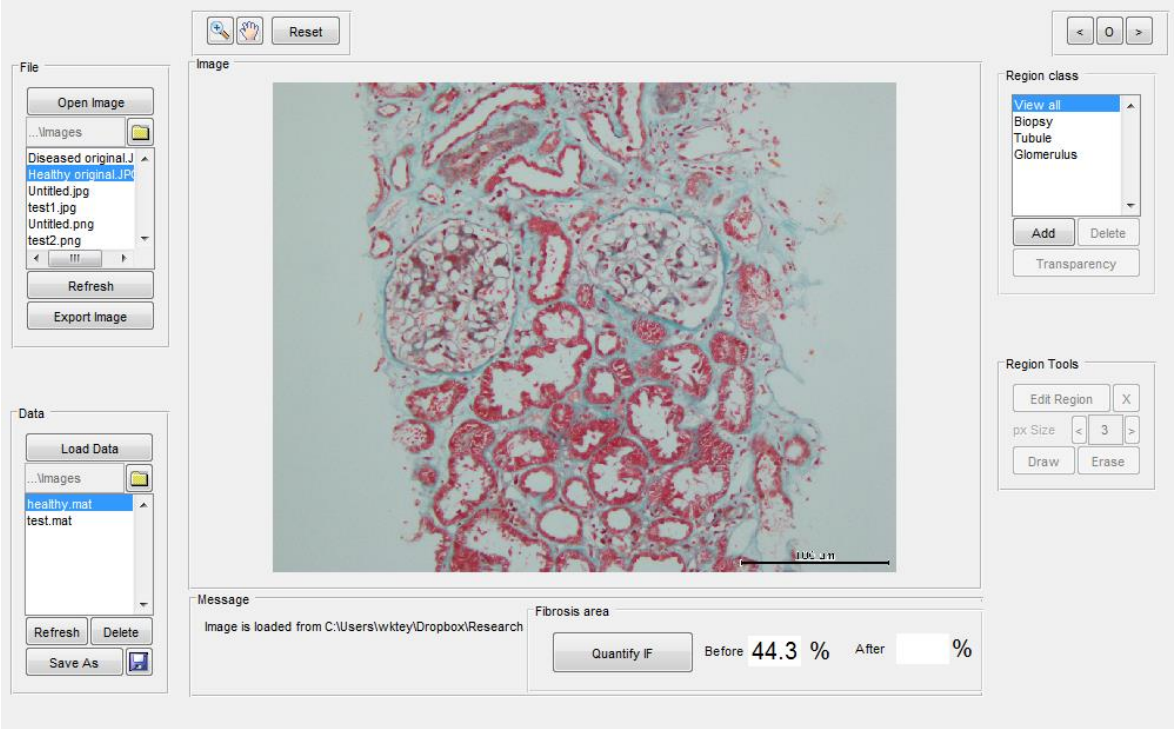


Figure 5.4 The prototype IF quantification GUI developed in the MATLAB environment. The GUI presents: the panel for selecting original biopsy images (**File panel**); the panel for loading IF quantification result data from the proposed quantification framework (**Data panel**); the panel for highlighting different tissue structure regions for visualisation of segmentation results (**Region class panel**); the panel for editing tissue structure regions by the user (**Region Tools panel**); the panel for the recalculation of IF area after the tissue structure region editing by the user (**Fibrosis area panel**).

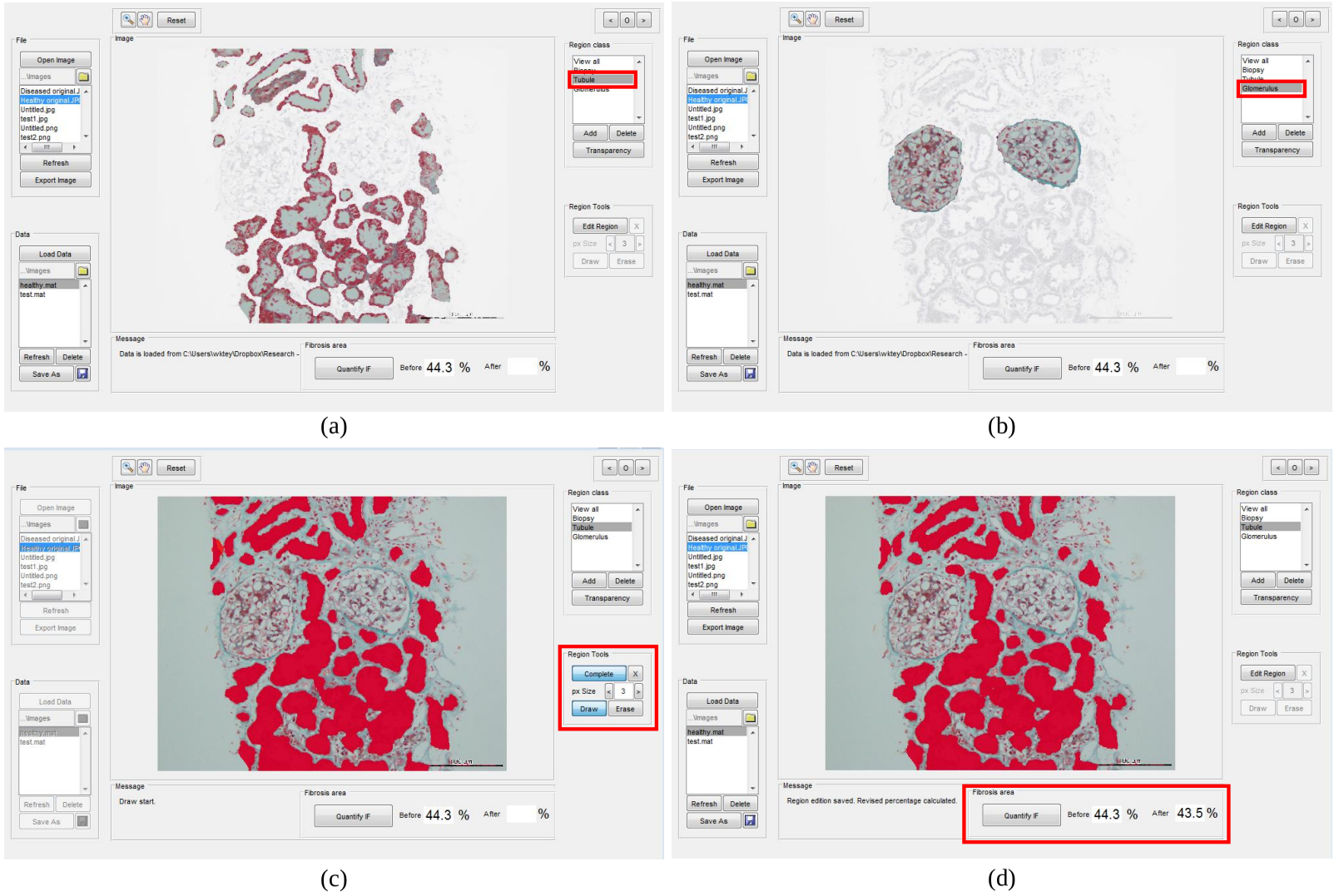


Figure 5.5 Operation of the GUI —Highlighting of the (a) renal tubular structures, and (b) glomeruli. (c) The editing tools for correcting automatically segmented tissue structure regions, and (d) the recalculated IF area in percentage.

Chapter 6

Conclusion

6.1 The current state of development

Computer aided diagnosis is a practice that is beginning to see more widespread implementation in hospitals and healthcare centres. This growth is necessitated by the advantages brought about by having a computerised aide in the diagnostic process. These include the reduced diagnosis time required, lowered need for laborious manual evaluation by medical personnel, as well as increased accuracy and objectivity in the medical analysis performed by doctors. The demand for computerised aides in medical diagnoses is further spurred by the advancement of computation technology and the improving affordability and accessibility of healthcare services. Besides, the explosion of digitised medical data also facilitates the introduction of computerised tools into medical diagnosis settings.

The applicability of computer aided diagnosis also extends to the diagnosis of kidney diseases. The kidney is an important blood filtering organ in the human circulatory system responsible for removing waste products from the blood. In the case of its failure, it can result in the accumulation of toxic waste in the body. This condition can be fatal

to the human body. Fortunately, chronic kidney disease can be controlled or slowed in its progression to end-stage renal failure through early diagnosis and treatment. Doctors and pathologists search for indications to a possible kidney disease through blood and urine tests in routine procedures. In definitive cases, a renal biopsy sample is taken from the patient as it is the gold standard to ascertaining the presence and activity of the kidney disease [3]. Among all the pathological structures to be found in the biopsy samples, the identifying structure that is indicative of chronic kidney disease is the IF.

Assessment of IF in biopsy samples is traditionally carried out through visual evaluation of the pathologists. This standard practice in the diagnosis process [6] requires the pathologists to approximate the total IF area in the biopsy sample cortical region by eye. The concern in this scenario lies in the absence of a universal yardstick for the pathologists to compare against. One measure that stands in as the closest to a standard for the pathologists is the Banff classification criteria for renal allograft diagnoses [8]. It provides a guideline for expressing the estimated amount of IF as a percentage of the biopsy tissue area in three grades: mild, moderate and severe. Unfortunately, this standard is still reliant on the judgement of the human observer in the estimation of IF region. To address this problem, a framework for the quantification of IF areas in digital biopsy images based on computer vision and machine learning has to be developed and applied with concordance to the pathologists' standard.

6.2 Research summary

Setting out to resolve the research problem, the state-of-the-art methods for automated renal IF quantification were investigated. These approaches can be summarised into two methods, namely colour thresholding and colour clustering. Both of these methods were replicated on the dataset compiled from biopsy images collected from the Sultanah Aminah Hospital as a baseline for comparison. On the other hand, a novel approach

based on the identification and segmentation of all commonly observed renal tissue structures was proposed as the framework for renal IF quantification in this project.

The conventional method of extracting the IF structure by utilising the colour information in the biopsy images entails the investigation into the different colour spaces presented by the digital biopsy colour images. The various colour spaces analysed were the *RGB*, *YCbCr*, *CIE Lab*, and *HSI* colour spaces. Segmentation of the IF areas by intensity thresholding and clustering in the colour space components corresponding to the blue-green colour – the colour of fibrosis structures as stained by the chemical dye, were executed. Experiments were carried out on a ground truth dataset consisting of 20 images of biopsy sections which only contains tubular structures stained by MT.

The main contribution of this research project is the proposition of a framework for IF quantification by the identification and extraction of all renal tissue structures. Each tissue structure is plainly defined in terms of computer vision rules and parameters to help in their segmentation. Inspired by the conventional method of utilising colour spaces, it is proposed that a selective combination of different colour spaces be exploited in the detection of red pixels and subsequently red structures in the image. This approach is used in conjunction with morphological operations to segment the red tubular structures in the biopsy images. By deduction, the IF structures can be identified as regions that are in the background of the tubular structures in the biopsy sample. A preliminary experiment on a dataset of images containing only tubular structure sections was carried out. The proposed tubular structure segmentation method for IF quantification was contrasted against the conventional colour thresholding and clustering methods, as previously mentioned. The findings from the experiment have proven that the quantification by detection of tubular structures achieved an average error half of that produced by the colour thresholding and clustering approaches. The feasibility of the IF quantification approach by the detection of all tissue structures has thus been validated.

Among all the renal tissue structures to be identified, the segmentation of healthy glomerulus structure is one of most challenging. The main reason for it still being an open research question is in the highly irregular appearance of the glomerulus in biopsy images. The glomerulus is a spherical tissue structure with fluid-filled spaces and can thus present large variations when being prepared into a 2D biopsy sample. This is dependent on the biopsy slicing orientation, conditions and the degree of glomerular atrophy. Three distinct techniques were introduced for the detection and segmentation of glomerulus structures: Bowman's space search, intraglomerular texture classification, and classification by a novel image texture feature named channel correlation feature.

The Bowman's space search and intraglomerular texture classification methods are variations of methods already tested in the literature. The search for the characteristic Bowman's space around the glomerulus is based on the premise of the detection of elliptical white spaces around low intensity intraglomerular regions. Meanwhile, the intraglomerular texture classification approach utilises image textural features such as GLCM, GLRL, SFTA, and colour histogram features to train an rLADTree classifier that detects glomerulus structures. On the other hand, the segmentation method by utilising the novel channel correlation feature as classification descriptor presents a more robust way of glomerulus detection than the other two methods. It is free from the rigid Bowman's shape constraint while is able to extract a higher layer of information than basic textural features. These two advantages have provided an edge to the classification by novel channel correlation feature method compared to the other methods in glomerulus segmentation, as proven in validation experiments.

The novel channel correlation feature generation method is founded on the measurement of the amount of correlation between different representations or interpretations of an image to be classified. This allows the creation of a feature that is scale, rotation and translation invariant. As opposed to common image textural features, the proposed channel correlation texture feature is not restricted to the measure of local homogeneity

but also maps the spatial characteristics of the image pixels. Validation experiments on comparing between the proposed feature and other state-of-the-art textural features have proven its superiority in classifying accuracy and knowledge abstraction. An in-depth assessment on the knowledge abstraction capability between the proposed textural feature and the CNN features have shown that the proposed feature offers a more interpretable feature set than the CNN features.

Finally, the impact of the finalised automated IF quantification system on the diagnostic process conducted by the pathologists was investigated. As a benchmark procedure, the variability of practising pathologists in the visual evaluation of renal IF in biopsy images was firstly measured and inspected. In a subsequent stage, they were equipped with the automated quantification of the amount of IF present in the biopsy image inspected produced by the proposed quantification system. Findings from the experiments have demonstrated that the proposed system substantially improved the pathologists' agreement in diagnosis. An interesting point to note is that the computerised quantification has exhibited a higher agreement to the average pathologist than the agreement shown between independent pathologists. This highlights the extent to which the pathologists can benefit, and have benefited, from having a precise computerised quantification system as an aide in their kidney disease diagnosis procedure.

6.3 Future research

The proposed system for the quantification of renal IF in digital kidney biopsy images by the method of tissue structure segmentation has produced satisfactory results based on the comparison with the pathologists' evaluation. However, further refinements of the system are still desired to improve the detection and recognition of individual structures. This problem is especially defined in the recognition of the healthy glomerulus structures. The current method for segmentation requires the calculation of different statistical

properties on top of the derivations of various image representations of the original image. This is a computationally feasible process in this research project, yet a simplification of the process is still favourable for more large scale applications. Proposed solutions to this include searching for more descriptive image representations for the dataset and the development of the program in a more efficient and faster programming language, such as C or Java, than the MATLAB environment.

Aside from that, the utilisation of SVM in the classification of glomerulus structure features can be replaced by an unsupervised or semi-supervised learning algorithm, such as the increasingly popular neural network in the scope of medical imaging. An adaptation of an unsupervised learning algorithm into glomerulus structure detection would remove the need of searching for an optimised set of parameters that can best generalise the dataset model. Additionally, examination of the layers in deep neural networks trained on the glomerulus dataset used in this research could also potentially provide insights to the way the image dataset is generalised.

Finally, in order to create a convincing validation of the quantification system's algorithm, an expansion of the digital biopsy image dataset has to be carried out by further image acquisition from the hospital. A larger dataset will introduce a broader range of variations seen in the biopsy images, thus bringing to light any possible limitations of the system previously unnoticed for further enhancement of the algorithm. Alongside the expanded image dataset, the set of manually annotated ground truth images will have to be increased as well to test for the accuracy of the system in a more comprehensive manner. Feedback from the pathologists can also be efficiently gathered through the developed GUI as the visual evaluation results are gathered for the new image dataset.

Appendix A

Compiled testing dataset

From the large image dataset obtained from the hospital, a smaller subset of renal biopsy images is selected as a testing dataset for the proposed experiments. The selection criterion for this compilation of biopsy images is based on the type of tissue structure observed in the section of biopsy sample in order to facilitate the execution of successive experiments. All the tissue structures commonly observed in a renal tissue is introduced and described in Chapter 1.1. Furthermore, the testing dataset was also assembled with the intention of encompassing all possible variation presented in the larger dataset.

The image datasets utilised in the experiments are listed as follows, with corresponding sample images in Figure A.1.

1. 140 images of biopsy consisting of only tubular structures.

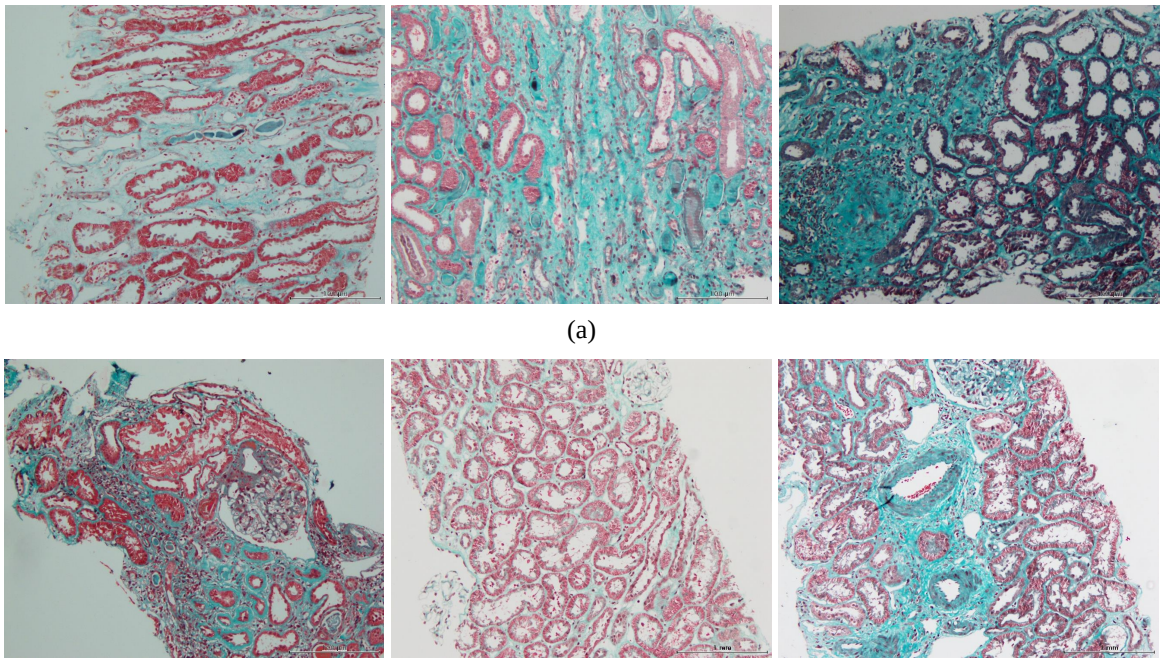
2. 146 images of biopsy with all possible tissue structures (renal tubules, glomeruli, arteries, etc.).

Additionally, for the experiment involving the validation of the novel feature descriptor based on channel correlation, a glomerulus classification dataset was also compiled from the existing image dataset. The dataset consists of:

1. 479 glomerulus images, and

2. 499 non-glomerulus images which comprise various surrounding tissue structures.

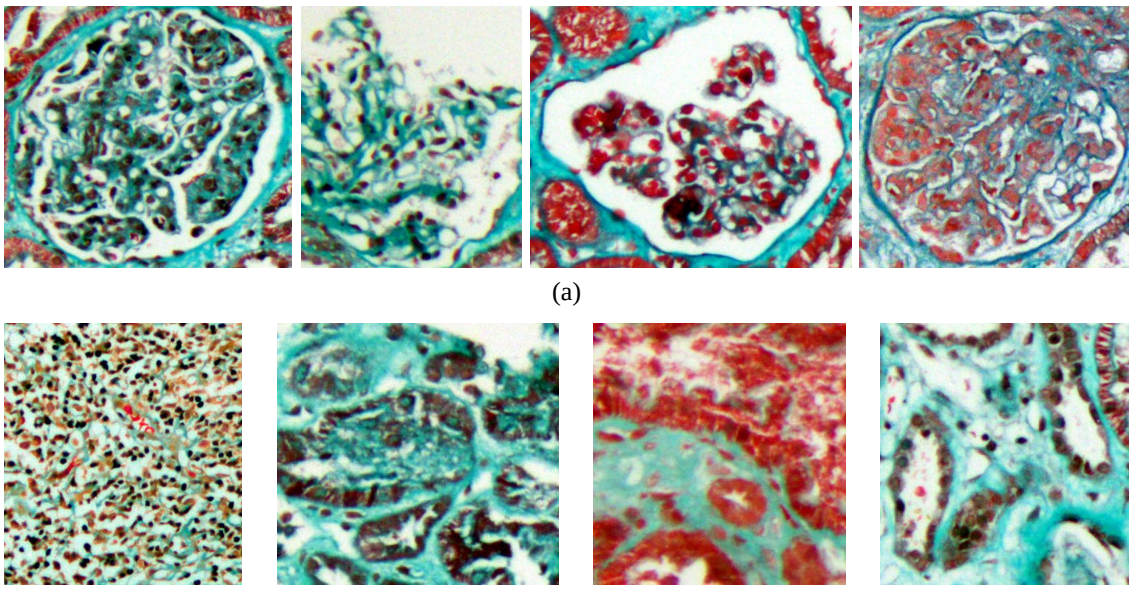
A few corresponding sample images are displayed in Figure A.2.



(a)

(b)

Figure A.1 Samples from the renal biopsy datasets with (a) only tubular structures, and (b) all commonly seen structures.



(a)

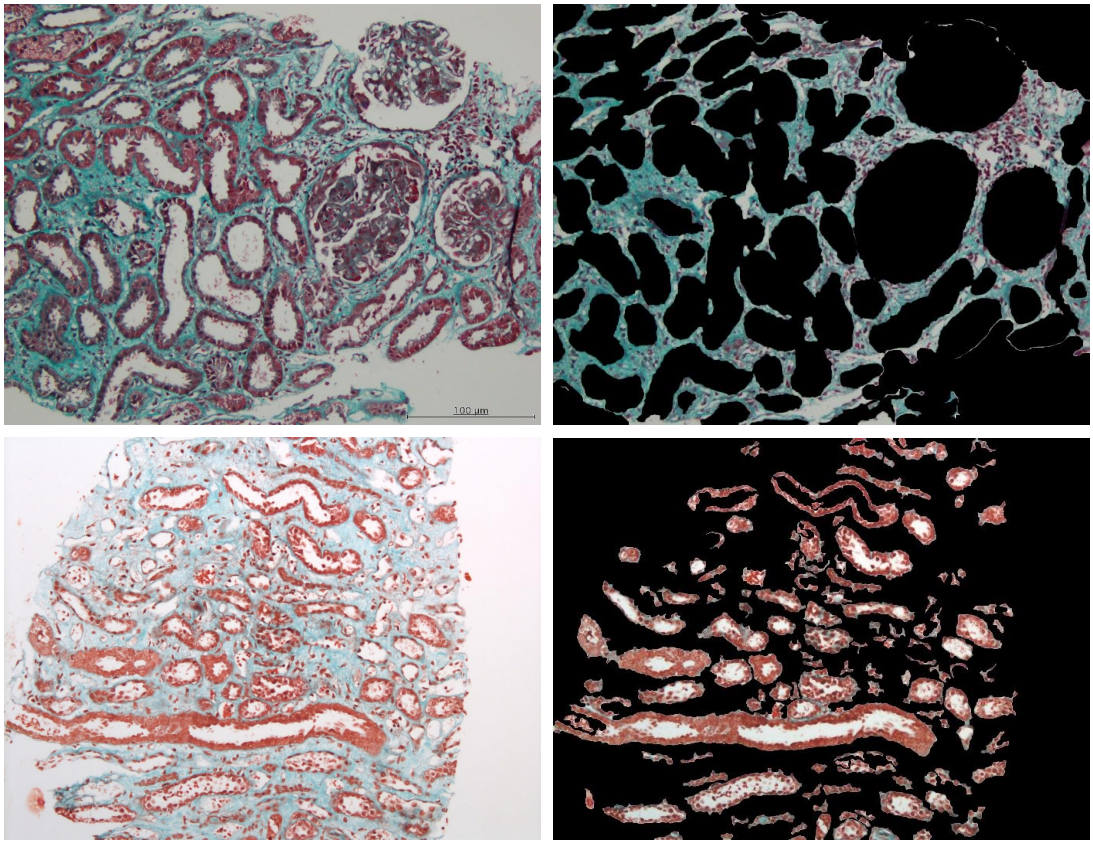
(b)

Figure A.2 Samples from the glomerulus classification dataset, with (a) glomerulus structures, and (b) various surrounding tissue structures.

Appendix B

Annotated ground truth

In order to provide an objective evaluation on the IF segmentation methods, a set of 40 images was selected from the bigger biopsy image dataset as the ground truth dataset. This dataset was selected to include 20 relatively healthy and 20 relatively diseased biopsies. Additionally, from each of the healthy and diseased set, 10 images contain only renal tubular structures while the other 10 include other common renal tissue structures as well. These images have their IF regions manually annotated by consulting an experienced pathologist. Examples from this image dataset is presented in Figure B.1.



(a) (b)

Figure B.1 (a) Sample renal biopsies. (b) Manually segmented ground truth images of IF regions.

Appendix C

Colour invariance

The colour invariant expressions are modelled under the assumption made by the Kubelka-Munk theory on the reflected light spectrum for coloured bodies [313]. The photometric reflectance model is stated as

$$E(\lambda,\bar{x}) = e(\lambda,\bar{x})(1-\rho_f(\bar{x}))^2 R_\infty(\lambda,\bar{x}) + e(\lambda,\bar{x})\rho_f(\bar{x}) \tag{C.1}$$

where λ is the wavelength,
$\bar{x}$ is the position in the image,
$e(\lambda,\bar{x})$ is the illumination spectrum,
$\rho_f(\bar{x})$ is the Fresnel reflectance, and
$R_\infty(\lambda,\bar{x})$ is the material reflectivity.

The distinct states of different colours are discriminated using colour invariance by adopting a few assumptions on the imaging conditions. For the purpose of inspecting biopsy images, matte, dull surfaces, planar objects, and an equal energy and uniform illumination in the Kubelka-Munk model are assumed. This reduces the illumination spectrum in Equation (C.1) to an intensity of i and the equation is simplified to

Equation (C.2). Differentiating Equation (C.2) with respect to x and normalising it results in Equation (C.3).

$$E(\lambda, x) = iR_{\infty}(\lambda, x) \tag{C.2}$$

$$W_x = \frac{E_x}{E} = \frac{1}{R_{\infty}(\lambda, x)} \frac{\partial R_{\infty}(\lambda, x)}{\partial x} \tag{C.3}$$

W_x in Equation (C.3) detects changes in object reflectance independent of illumination intensity. This property can also be used as an edge detector in coloured images. The complete and irreducible set of invariants for Equation (C.3) can be stated as Equation (C.4).

$$W_{\lambda^m x^n} = \frac{E_{\lambda^m x^n}}{E} \tag{C.4}$$

for any integers $m \geq 0$, $n \geq 1$. Therefore, the total edge strength measures in a two dimensional colour image may be defined as Equation (C.5).

$$W_W = \sqrt{{W_x}^2 + W_{\lambda x}^2 + W_{\lambda\lambda x}^2 + W_y^2 + W_{\lambda y}^2 + W_{\lambda\lambda y}^2} \tag{C.5}$$

This allows the colour invariance to function as an edge detector, where the edges present in the image based on colour gradient were derived independent of intensity level. In the context of renal biopsy foreground segmentation, the colour invariance edges were then morphologically extracted as a whole as the biopsy tissue region.

Appendix D

Statistical evaluation methods

Intraclass correlation coefficient

The ICC is used to measure the relative similarity between variables sharing the same metric and variance [314]. It is often applied as a measure of reliability of a group of ratings or measurements. The theoretical formula for the ICC [308] is given as

$$ICC = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_b^2} \tag{D.1}$$

where σ_r^2 is the variance of ratings for a subject, and σ_b^2 is the variance of ratings between subjects. In other words, the ICC is a measure of rating reliability by comparing the variability of ratings for the same subject compared to the total variability of all ratings and subjects.

However, the values for σ_r^2 and σ_b^2 are usually not known, and they are estimated from a sample data set. These estimates are dependent on the different models used for the sample data, so as to acquire the mean square values for analysis of variance (ANOVA) suitable for the sample data. The three types of ICC [315] are:

Case 1: For each subject, a rater is selected at random from a pool of raters to give a rating on the subject. This corresponds to a one-way random effects ANOVA model.

Case 2: For each subject, the same set of k raters selected at random from a pool of raters give ratings to the subject. This corresponds to a two-way random effects ANOVA model.

Case 3: For each subject, the same set of k raters give ratings to the subject, and there are only k raters. This corresponds to a two-way mixed effect ANOVA model.

Two ICCs exist for each of the two-way data cases, namely a correlation measuring the consistency of ratings, and a correlation of absolute agreement between ratings. The difference between these two is in the exclusion of raters' variance from the denominator (total) variance of ICC for consistency measures, while the absolute agreement measure retains the raters' variance. For instance, in the measure of consistency between raters, a Pathologist 1 giving relatively high percentage fibrosis scores may agree with a Pathologist 2 giving relatively low percentage fibrosis scores if their scores can be related by a linear shift in their mean scores.

Kappa coefficient

The kappa coefficient is a statistical tool for measuring the inter-rater reliability. Two types of kappa coefficient are commonly used in machine learning, namely the Cohen's kappa [316] and Fleiss' kappa [317]. These kappa coefficients are both used for assessing the agreement between a fixed number of raters on categorical ratings with chance agreement taken into account. However, the difference lies in the fact that the Cohen's kappa can only measure agreement between two raters, while the Fleiss' kappa is used in the case of more than two raters.

The Cohen's kappa coefficient, κ_C is calculated as

$$\kappa_C = \frac{p_o - p_e}{1 - p_e} \tag{D.2}$$

where p_o is the observed proportion of agreement between raters, and p_e is the calculated probability of chance agreement. Similarly, the Fleiss' kappa coefficient, κ_F has the equation of

$$\kappa_F = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{D.3}$$

where $\bar{P}$ is the mean proportion of agreement between all raters, and $\bar{P}_e$ is the calculated probability of chance agreement of all raters.

For both the kappa coefficients described, a complete agreement between the raters would produce $\kappa = 1$, while a value of $\kappa \leq 0$ would be obtained if there is no agreement between the raters other than what would be expected by chance.

References

[1] V. Jha, G. Garcia-Garcia, K. Iseki, Z. Li, S. Naicker, B. Plattner, R. Saran, A. Y.-M. Wang, and C.-W. Yang, "Chronic kidney disease: global dimension and perspectives," *The Lancet*, vol. 382, no. 9888, pp. 260–272, 2013.

[2] G. Zoppini, G. Targher, M. Chonchol, V. Ortalda, C. Negri, V. Stoico, and E. Bonora, "Predictors of estimated gfr decline in patients with type 2 diabetes and preserved kidney function," *Clinical Journal of the American Society of Nephrology*, pp. CJN–07 650 711, 2012.

[3] B. Afzali, S. Jayawardene, and D. Goldsmith, "Diagnostic tests in chronic kidney disease," *Kidney Disease*, p. 1, 2007.

[4] T. D. Hewitson, "Renal tubulointerstitial fibrosis: common but never simple," *American Journal of Physiology-Renal Physiology*, vol. 296, no. 6, pp. F1239–F1244, 2009.

[5] A. Bohle, S. Mackensen-Haen, and H. v. Gise, "Significance of tubulointerstitial changes in the renal cortex for the excretory function and concentration ability of the kidney: a morphometric contribution," *American Journal of Nephrology*, vol. 7, no. 6, pp. 421–433, 1987.

[6] F. Moreso, M. Lopez, A. Vallejos, C. Giordani, L. Riera, X. Fulladosa, M. Hueso, J. Alsina, J. M. Grinyó, and D. Serón, "Serial protocol biopsies to quantify the progression of chronic transplant nephropathy in stable renal allografts," *American Journal of Transplantation*, vol. 1, no. 1, pp. 82–88, 2001.

[7] J. M. Street, A. C. P. Souza, A. Alvarez-Prats, T. Horino, X. Hu, P. S. Yuen, and R. A. Star, "Automated quantification of renal fibrosis with sirius red and polarization contrast microscopy," *Physiological Reports*, vol. 2, no. 7, p. e12088, 2014.

[8] A. Loupy, M. Haas, K. Solez, L. Racusen, D. Glotz, D. Seron, B. Nankivell, R. Colvin, M. Afrouzian, E. Akalin *et al.*, "The banff 2015 kidney meeting report: current challenges in rejection classification and prospects for adopting molecular pathology," *American Journal of Transplantation*, vol. 17, no. 1, pp. 28–41, 2017.

[9] D. C. Cattran, R. Coppo, H. T. Cook, J. Feehally, I. S. Roberts, S. Troyanov, C. E. Alpers, A. Amore, J. Barratt, F. Berthoux *et al.*, "The oxford classification

of iga nephropathy: rationale, clinicopathological correlations, and classification," *Kidney International*, vol. 76, no. 5, pp. 534–545, 2009.

[10] C. Hsieh, A. Chang, D. Brandt, R. Guttikonda, T. O. Utset, and M. R. Clark, "Tubulointerstitial inflammation and scarring predict outcome in lupus nephritis," *Arthritis Care & Research*, vol. 63, no. 6, p. 865, 2011.

[11] N. Marcussen, T. Olsen, H. Benediktsson, L. Racusen, and K. Solez, "Reproducibility of the banff classification of renal allograft pathology. inter-and intraobserver variation." *Transplantation*, vol. 60, no. 10, pp. 1083–1089, 1995.

[12] P. Furness and N. Taub, "Convergence of european renal transplant pathology assessment procedures p. international variation in the interpretation of renal transplant biopsies: report of the certpap project," *Kidney International*, vol. 60, no. 5, pp. 1998–2012, 2001.

[13] D. Serón, F. Moreso, X. Fulladosa, M. Hueso, M. Carrera, and J. M. Grinyó, "Reliability of chronic allograft nephropathy diagnosis in sequential protocol biopsies," *Kidney International*, vol. 61, no. 2, pp. 727–733, 2002.

[14] T. J. Fuchs and J. M. Buhmann, "Computational pathology: Challenges and promises for tissue analysis," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 515–530, 2011.

[15] V. Meas-Yedid, A. Servais, L.-H. Noël, C. Panterne, P. Landais, N. Hervé, N. Brousse, H. Kreis, C. Legendre, E. Thervet *et al.*, "New computerized color image analysis for the quantification of interstitial fibrosis in renal transplantation," *Transplantation*, vol. 92, no. 8, pp. 890–899, 2011.

[16] R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 224–235, 2007.

[17] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[18] ——, "Current status and future potential of computer-aided diagnosis in medical imaging," *The British Journal of Radiology*, vol. 78, no. suppl_1, pp. s3–s19, 2005.

[19] A. Slater, S. A. Taylor, E. Tam, L. Gartner, J. Scarth, C. Peiris, A. Gupta, M. Marshall, D. Burling, and S. Halligan, "Reader error during ct colonography: causes and implications for training," *European Radiology*, vol. 16, no. 10, pp. 2275–2283, 2006.

[20] R. Takahashi and Y. Kajikawa, "Computer-aided diagnosis: a survey with bibliometric analysis," *International Journal of Medical Informatics*, vol. 101, pp. 58–67, 2017.

[21] F. Winsberg, M. Elkin, J. Macy Jr, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, no. 2, pp. 211–215, 1967.

[22] F. Roellinger Jr, A. Kahveci, J. Chang, C. A. Harlow, S. J. Dwyer III, and G. S. Lodwick, "Computer analysis of chest radiographs," *Computer Graphics and Image Processing*, vol. 2, no. 3-4, pp. 232–251, 1973.

[23] H. Becker, W. Nettleton, P. Meyers, J. Sweeney, and C. Nice, "Digital computer determination of a medical diagnostic index directly from chest x-ray images," *IEEE Transactions on Biomedical Engineering*, no. 3, pp. 67–72, 1964.

[24] E. A. Sickles, "Breast calcifications: mammographic evaluation." *Radiology*, vol. 160, no. 2, pp. 289–293, 1986.

[25] K. J. McLoughlin, P. J. Bones, N. Karssemeijer *et al.*, "Noise equalization for detection of microcalcification clusters in direct digital mammogram images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 3, pp. 313–320, 2004.

[26] J.-J. Mordang, A. Gubern-Mérida, G. den Heeten, and N. Karssemeijer, "Reducing false positives of microcalcification detection systems by removal of breast arterial calcifications," *Medical Physics*, vol. 43, no. 4, pp. 1676–1687, 2016.

[27] X. Liu, M. Mei, J. Liu, and W. Hu, "Microcalcification detection in full-field digital mammograms with pfcm clustering and weighted svm-based method," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 73, 2015.

[28] P. Casaseca-de-la Higuera, J. I. Arribas, E. Munoz-Moreno, and C. Alberola-López, "A comparative study on microcalcification detection methods with posterior probability estimation based on gaussian mixture models," in *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005.* IEEE, 2006, pp. 49–54.

[29] G. Lemaur, K. Drouiche, and J. DeConinck, "Highly regular wavelets for the detection of clustered microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 393–401, 2003.

[30] J. Dheeba and S. T. Selvi, "A swarm optimized neural network system for classification of microcalcification in mammograms," *Journal of Medical Systems*, vol. 36, no. 5, pp. 3051–3061, 2012.

[31] R. M. Rangayyan, N. M. El-Faramawy, J. L. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 799–810, 1997.

[32] W. P. Kegelmeyer Jr, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions." *Radiology*, vol. 191, no. 2, pp. 331–337, 1994.

[33] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli, "A novel featureless approach to mass detection in digital mammograms based on support vector machines," *Physics in Medicine & Biology*, vol. 49, no. 6, p. 961, 2004.

[34] H.-P. Chan, D. Wei, M. A. Helvie, B. Sahiner, D. D. Adler, M. M. Goodsitt, and N. Petrick, "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space," *Physics in Medicine & Biology*, vol. 40, no. 5, p. 857, 1995.

[35] R. Bellotti, F. De Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Fauci, R. Magro *et al.*, "A completely automated cad system for mass detection in a large mammographic database," *Medical Physics*, vol. 33, no. 8, pp. 3066–3075, 2006.

[36] A. C. of Radiology. BI-RADS Committee, *Breast imaging reporting and data system*. American College of Radiology, 1998.

[37] R. M. Rangayyan and F. J. Ayres, "Gabor filters and phase portraits for the detection of architectural distortion in mammograms," *Medical and Biological Engineering and Computing*, vol. 44, no. 10, pp. 883–894, 2006.

[38] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1215–1227, Dec 2001.

[39] T. Matsubara, T. Ichikawa, T. Hara, H. Fujita, S. Kasai, T. Endo, and T. Iwase, "Automated detection methods for architectural distortions around skinline and within mammary gland on mammograms," in *International Congress Series*, vol. 1256. Elsevier, 2003, pp. 950–955.

[40] M. J. Homer and E. A. Sickles, *Mammographic interpretation: a practical approach*. McGraw-Hill New York, 1997, vol. 224.

[41] P. Miller and S. Astley, "Automated detection of mammographic asymmetry using anatomical features," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 06, pp. 1461–1476, 1993.

[42] ——, "Automated detection of breast asymmetry using anatomical features," in *State of the Art in Digital Mammographic Image Analysis*. World Scientific, 1994, pp. 247–261.

[43] R. J. Ferrari, R. M. Rangayyan, J. L. Desautels, and A. F. Frère, "Analysis of asymmetry in mammograms via directional filtering with gabor wavelets," *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 953–964, 2001.

[44] C. E. Ravin and H. G. Chotas, "Chest radiography." *Radiology*, vol. 204, no. 3, pp. 593–600, 1997.

[45] S. Kollins, "Computed tomography of the pulmonary parenchyma and chest wall." *Radiologic Clinics of North America*, vol. 15, no. 3, pp. 297–308, 1977.

[46] V. J. Lowe and K. S. Naunheim, "Current role of positron emission tomography in thoracic oncology," *Thorax*, vol. 53, no. 8, pp. 703–712, 1998.

[47] S. Diederich, M. Lentschig, T. Overbeck, D. Wormanns, and W. Heindel, "Detection of pulmonary nodules at spiral ct: comparison of maximum intensity projection sliding slabs and single-image reporting," *European Radiology*, vol. 11, no. 8, pp. 1345–1350, 2001.

[48] N. Nakamori, K. Doi, H. Macmahon, Y. Sasaki, and S. Montner, "Effect of heart-size parameters computed from digital chest radiographs on detection of cardiomegaly. potential usefulness for computer-aided diagnosis." *Investigative Radiology*, vol. 26, no. 6, pp. 546–550, 1991.

[49] K. Abe, K. Doi, H. MacMahon, M. L. Giger, H. Jia, X. Chen, A. Kano, and T. Yanagisawa, "Computer-aided diagnosis in chest radiography. preliminary experience." *Investigative Radiology*, vol. 28, no. 11, pp. 987–993, 1993.

[50] S. Jaeger, A. Karargyris, S. Candemir, J. Siegelman, L. Folio, S. Antani, and G. Thoma, "Automatic screening for tuberculosis in chest radiographs: a survey," *Quantitative Imaging in Medicine and Surgery*, vol. 3, no. 2, p. 89, 2013.

[51] K. Suzuki, M. Kusumoto, S.-i. Watanabe, R. Tsuchiya, and H. Asamura, "Radiologic classification of small adenocarcinoma of the lung: radiologic-pathologic correlation and its prognostic impact," *The Annals of Thoracic Surgery*, vol. 81, no. 2, pp. 413–419, 2006.

[52] H. Chen, Y. Xu, Y. Ma, and B. Ma, "Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on ct images: clinical evaluation," *Academic Radiology*, vol. 17, no. 5, pp. 595–602, 2010.

[53] S. Soltaninejad, M. Keshani, and F. Tajeripour, "Lung nodule detection by knn classifier and active contour modelling and 3d visualization," in *16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), 2012*. IEEE, 2012, pp. 440–445.

[54] A. Riccardi, T. S. Petkov, G. Ferri, M. Masotti, and R. Campanini, "Computer-aided detection of lung nodules via 3d fast radial transform, scale space representation, and zernike mip classification," *Medical Physics*, vol. 38, no. 4, pp. 1962–1971, 2011.

[55] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, C.-T. Zhang, K. Suzuki *et al.*, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Medical Physics*, vol. 30, no. 9, pp. 2440–2454, 2003.

[56] H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz *et al.*, "Computer-aided diagnosis in chest radiography: results of large-scale observer tests at the 1996–2001 rsna scientific assemblies," *Radiographics*, vol. 23, no. 1, pp. 255–265, 2003.

[57] A. Teuner, O. Pichler, and B. J. Hosticka, "Unsupervised texture segmentation of images using tuned matched gabor filters," *IEEE Transactions on Image Processing*, vol. 4, no. 6, pp. 863–870, 1995.

[58] B. Van Ginneken, B. T. H. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.

[59] C.-M. Pun and M.-C. Lee, "Log-polar wavelet energy signatures for rotation and scale invariant texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 590–603, 2003.

[60] J. Y. Hong, Z. Bin, S. Berkman, and C. D. P., "Evaluating computer-aided detection algorithms," *Medical Physics*, vol. 34, no. 6Part1, pp. 2024–2038. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2736289

[61] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[62] S. Winawer, R. Fletcher, D. Rex, J. Bond, R. Burt, J. Ferrucci, T. Ganiats, T. Levin, S. Woolf, D. Johnson *et al.*, "Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence," *Gastroenterology*, vol. 124, no. 2, pp. 544–560, 2003.

[63] A. H. Dachman and A. Laghi, *Atlas of virtual colonoscopy*. Springer Science & Business Media, 2010.

[64] J. Näppi and H. Yoshida, "Automated detection of polyps with ct colonography: evaluation of volumetric features for reduction of false-positive findings," *Academic Radiology*, vol. 9, no. 4, pp. 386–397, 2002.

[65] H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1261–1274, 2001.

[66] S. Wang, J. Yao, and R. M. Summers, "Improved classifier for computer-aided polyp detection in ct colonography by nonlinear dimensionality reduction," *Medical Physics*, vol. 35, no. 4, pp. 1377–1386, 2008.

[67] J. Li, R. Van Uitert, J. Yao, N. Petrick, M. Franaszek, A. Huang, and R. M. Summers, "Wavelet method for ct colonography computer-aided polyp detection," *Medical Physics*, vol. 35, no. 8, pp. 3527–3538, 2008.

[68] Z. Wang, Z. Liang, L. Li, X. Li, B. Li, J. Anderson, and D. Harrington, "Reduction of false positives by internal features for polyp detection in ct-based virtual colonoscopy," *Medical Physics*, vol. 32, no. 12, pp. 3602–3616, 2005.

[69] K. Suzuki, D. C. Rockey, and A. H. Dachman, "Ct colonography: Advanced computer-aided detection scheme utilizing mtanns for detection of "missed" polyps in a multicenter clinical trial," *Medical Physics*, vol. 37, no. 1, pp. 12–21, 2010.

[70] K. Suzuki, H. Yoshida, J. Näppi, S. G. Armato, and A. H. Dachman, "Mixture of expert 3d massive-training anns for reduction of multiple types of false positives in cad for detection of polyps in ct colonography," *Medical Physics*, vol. 35, no. 2, pp. 694–703, 2008.

[71] H. Yoshida and A. Dachman, "Cad techniques, challenges, andcontroversies in computed tomographic colonography," *Abdominal Imaging*, vol. 30, no. 1, pp. 26–41, 2004.

[72] M. Morrin, J. Sosna, J. Kruskal, A. Blachard, N. Stanietzky, and B. Siewert, "Diagnostic performance of radiologists with differing levels of expertise in the evaluation of ct colonography," *Radiology*, vol. 226, no. P, p. 365, 2003.

[73] J. J. Näppi, H. Frimmel, A. H. Dachman, and H. Yoshida, "Computerized detection of colorectal masses in ct colonography based on fuzzy merging and wall-thickening analysis," *Medical Physics*, vol. 31, no. 4, pp. 860–872, 2004.

[74] J. Feng, N. G. Isern, S. D. Burton, and J. Z. Hu, "Studies of secondary melanoma on c57bl/6j mouse liver using 1h nmr metabolomics," *Metabolites*, vol. 3, no. 4, pp. 1011–1035, 2013.

[75] H. L. Kaufman, *The melanoma book: a complete guide to prevention and treatment*. Gotham Books, 2005.

[76] E. Unlu, B. N. Akay, and C. Erdem, "Comparison of dermatoscopic diagnostic algorithms based on calculation: The abcd rule of dermatoscopy, the seven-point checklist, the three-point checklist and the cash algorithm in dermatoscopic evaluation of melanocytic lesions," *The Journal of Dermatology*, vol. 41, no. 7, pp. 598–603, 2014.

[77] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, and D. Polsky, "Early diagnosis of cutaneous melanoma: revisiting the abcd criteria," *Journal of the American Medical Association*, vol. 292, no. 22, pp. 2771–2776, 2004.

[78] R. Mackie and V. Doherty, "Seven-point checklist for melanoma," *Clinical and Experimental Dermatology*, vol. 16, no. 2, pp. 151–152, 1991.

[79] I. Zalaudek, G. Argenziano, H. Soyer, R. Corona, F. Sera, A. Blum, R. Braun, H. Cabo, G. Ferrara, A. Kopf *et al.*, "Three-point checklist of dermoscopy: an open internet study," *British Journal of Dermatology*, vol. 154, no. 3, pp. 431–437, 2006.

[80] S. W. Menzies, K. A. Crotty, C. Ingvar, and W. H. McCarthy, *An atlas of surface microscopy of pigmented skin lesions: dermoscopy*. McGraw-Hill Sydney, Australia, 2003.

[81] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," *Dermoscopy Image Analysis*, pp. 97–129, 2015.

[82] Q. Abbas, I. F. Garcia, M. Emre Celebi, W. Ahmad, and Q. Mushtaq, "A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images," *Skin Research and Technology*, vol. 19, no. 1, pp. e490–e497, 2013.

[83] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.

[84] K. Kiani and A. R. Sharafat, "E-shaver: An improved dullrazor for digitally removing dark and light-colored hairs in dermoscopic images," *Computers in Biology and Medicine*, vol. 41, no. 3, pp. 139–145, 2011.

[85] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[86] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines, C. Yu, A. Dhawan, and A. Huntley, "Segmentation of skin cancer images," *Image and Vision Computing*, vol. 17, no. 1, pp. 65–74, 1999.

[87] D. Ruiz, V. Berenguer, A. Soriano, and B. SáNchez, "A decision support system for the diagnosis of melanoma: A comparative approach," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15 217–15 223, 2011.

[88] B. Peng, L. Zhang, and D. Zhang, "Automatic image segmentation by dynamic region merging," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3592–3605, 2011.

[89] M. Emre Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz *et al.*, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.

[90] Z. Liu and J. Zerubia, "Skin image illumination modeling and chromophore identification for melanoma diagnosis," *Physics in Medicine & Biology*, vol. 60, no. 9, p. 3415, 2015.

[91] M. Sadeghi, M. Razmara, T. K. Lee, and M. S. Atkins, "A novel method for detection of pigment network in dermoscopic images using graphs," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 137–143, 2011.

[92] P. Schmid, "Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding," in *1999 International Conference on Image Processing Proceedings, 1999*, vol. 3. IEEE, 1999, pp. 449–453.

[93] S. O. Skrøvseth, T. R. Schopf, K. Thon, M. Zortea, M. Geilhufe, K. Møllersen, H. M. Kirchesch, and F. Godtliebsen, "A computer aided diagnostic system for malignant melanomas," in *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010.* IEEE, 2010, pp. 1–5.

[94] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 148–153, 2009.

[95] M. Zortea, E. Flores, and J. Scharcanski, "A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images," *Pattern Recognition*, vol. 64, pp. 92–104, 2017.

[96] R. B. Oliveira, N. Marranghello, A. S. Pereira, and J. M. R. Tavares, "A computational approach for detecting pigmented skin lesions in macroscopic images," *Expert Systems with Applications*, vol. 61, pp. 53–63, 2016.

[97] Q. Abbas, M. E. Celebi, and I. Garcia, "A novel perceptually-oriented approach for skin tumor segmentation," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 3, pp. 1837–1848, 2012.

[98] A. Sáez, B. Acha, and C. Serrano, "Pattern analysis in dermoscopic images," in *Computer Vision Techniques for the Diagnosis of Skin Cancer.* Springer, 2014, pp. 23–48.

[99] W. V. Stoecker, M. Wronkiewiecz, R. Chowdhury, R. J. Stanley, J. Xu, A. Bangert, B. Shrestha, D. A. Calcara, H. S. Rabinovitz, M. Oliviero *et al.*, "Detection of granularity in dermoscopy images of malignant melanoma using color and texture features," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 144–147, 2011.

[100] M. A. H. Bhuiyan, I. Azad, and K. Uddin, "Image processing for skin cancer features extraction," *International Journal of Scientific & Engineering Research*, vol. 4, no. 2, pp. 1–6, 2013.

[101] J. F. Alcón, C. Ciuhu, W. Ten Kate, A. Heinrich, N. Uzunbajakava, G. Krekels, D. Siem, and G. De Haan, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 14–25, 2009.

[102] A. R. Patel, V. Vejjabhinanta, and K. Nouri, "Clinical pearl: the evaluation of the surface area of small pigmented lesions," *International Journal of Dermatology*, vol. 46, no. 8, pp. 872–874, 2007.

[103] J. L. G. Arroyo and B. G. Zapirain, "Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis," *Computers in Biology and Medicine*, vol. 44, pp. 144–157, 2014.

[104] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, "Non-melanoma skin lesion classification using colour image data in a hierarchical k-nn classifier," in *9th IEEE International Symposium on Biomedical Imaging (ISBI), 2012*. IEEE, 2012, pp. 358–361.

[105] B. Shrestha, J. Bishop, K. Kam, X. Chen, R. H. Moss, W. V. Stoecker, S. Umbaugh, R. J. Stanley, M. E. Celebi, A. A. Marghoob *et al.*, "Detection of atypical texture features in early malignant melanoma," *Skin Research and Technology*, vol. 16, no. 1, pp. 60–65, 2010.

[106] R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule," *IET Image Processing*, vol. 10, no. 6, pp. 448–455, 2016.

[107] M. Lingala, R. J. Stanley, R. K. Rader, J. Hagerty, H. S. Rabinovitz, M. Oliviero, I. Choudhry, and W. V. Stoecker, "Fuzzy logic color detection: Blue areas in melanoma dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 38, no. 5, pp. 403–410, 2014.

[108] I. Maglogiannis and C. N. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721–733, 2009.

[109] D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri, and S. Chimenti, "Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: a comparative study," *British Journal of Dermatology*, vol. 147, no. 3, pp. 481–486, 2002.

[110] A. Mehta, A. S. Parihar, and N. Mehta, "Supervised classification of dermoscopic images using optimized fuzzy clustering based multi-layer feed-forward neural network," in *International Conference on Computer, Communication and Control (IC4), 2015*. IEEE, 2015, pp. 1–6.

[111] M. Zorman, M. M. Štiglic, P. Kokol, and I. Malčić, "The limitations of decision trees and automatic learning in real world medical decision making," *Journal of Medical Systems*, vol. 21, no. 6, pp. 403–415, 1997.

[112] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 670–677, 2008.

[113] Y. Zhou, M. Smith, L. Smith, A. Farooq, and R. Warr, "Enhanced 3d curvature pattern and melanoma diagnosis," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 155–165, 2011.

[114] J. Premaladha and K. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of Medical Systems*, vol. 40, no. 4, p. 96, 2016.

[115] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, "The 2007 who classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 2, pp. 97–109, 2007.

[116] L. P. Clarke, R. P. Velthuizen, M. Clark, J. Gaviria, L. Hall, D. Goldgof, R. Murtagh, S. Phuphanich, and S. Brem, "Mri measurement of brain tumor response: comparison of visual metric and automatic segmentation," *Magnetic Resonance Imaging*, vol. 16, no. 3, pp. 271–279, 1998.

[117] G. L. Qurat-Ul-Ain, S. B. Kazmi, M. A. Jaffar, and A. M. Mirza, "Classification and segmentation of brain tumor using texture analysis," *Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases*, pp. 147–155, 2010.

[118] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through mri: A survey and a new algorithm," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.

[119] S. Roy, S. Nag, I. K. Maitra, and S. K. Bandyopadhyay, "A review on automated brain tumor detection and segmentation from mri of brain," *arXiv preprint arXiv:1312.6150*, 2013.

[120] A. Amruta, A. Gole, and Y. Karunakar, "A systematic algorithm for 3-d reconstruction of mri based brain tumors using morphological operators and bicubic interpolation," in *2nd International Conference on Computer Technology and Development (ICCTD), 2010*. IEEE, 2010, pp. 305–309.

[121] M. Shokouhifar and G. S. Abkenar, "An artificial bee colony optimization for mri fuzzy segmentation of brain tissue," in *2011 International Conference on Management and Artificial Intelligence IPEDR*, vol. 6, 2011.

[122] N. Zhang, S. Ruan, S. Lebonvallet, Q. Liao, and Y. Zhu, "Kernel feature selection to fuse multi-spectral mri images for brain tumor segmentation," *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 256–269, 2011.

[123] M. F. Zarandi, M. Zarinbal, and M. Izadi, "Systematic image processing for diagnosing brain tumors: A type-ii fuzzy expert system approach," *Applied Soft Computing*, vol. 11, no. 1, pp. 285–294, 2011.

[124] S. Herlidou-Meme, J. Constans, B. Carsin, D. Olivie, P. Eliat, L. Nadal-Desbarats, C. Gondry, E. Le Rumeur, I. Idy-Peretti, and J. De Certaines, "Mri texture analysis on texture test objects, normal brain and intracranial tumors," *Magnetic Resonance Imaging*, vol. 21, no. 9, pp. 989–993, 2003.

[125] M. Maitra and A. Chatterjee, "Hybrid multiresolution slantlet transform and fuzzy c-means clustering approach for normal-pathological brain mr image segregation," *Medical Engineering & Physics*, vol. 30, no. 5, pp. 615–623, 2008.

[126] K. Gasmi, A. Kharrat, M. B. Messaoud, and M. Abid, "Automated segmentation of brain tumor using optimal texture features and support vector machine classifier,"

in *International Conference Image Analysis and Recognition*. Springer, 2012, pp. 230–239.

[127] M. A. Kassimi *et al.*, "3d model retrieval based on semantic and shape indexes," *arXiv preprint arXiv:1111.6387*, 2011.

[128] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[129] E. F. Badran, E. G. Mahmoud, and N. Hamdy, "An algorithm for detecting brain tumors in mri images," in *International Conference on Computer Engineering and Systems (ICCES), 2010*. IEEE, 2010, pp. 368–373.

[130] N. Pradhan and A. Sinha, "Fuzzy ann based detection and analysis of pathological and healthy tissues in flair magnetic resonance images of brain," *International Journal of Information Technology and Knowledge Management*, vol. 4, no. 2, pp. 471–476, 2011.

[131] S. Chaplot, L. Patnaik, and N. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86–92, 2006.

[132] K. Selvanayaki and P. Kalugasalam, "Performance evaluation of genetic algorithm segmentation on magnetic resonance images for detection of brain tumor," *European Journal of Scientific Research, ISSN*, pp. 365–378, 2012.

[133] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane, and M. Abid, "A hybrid approach for automatic classification of brain mri using genetic algorithm and support vector machine," *Leonardo Journal of Sciences*, vol. 17, no. 1, pp. 71–82, 2010.

[134] T. D. Wager and T. E. Nichols, "Optimization of experimental design in fmri: a general framework using a genetic algorithm," *Neuroimage*, vol. 18, no. 2, pp. 293–309, 2003.

[135] M. P. Arakeri and G. R. M. Reddy, "Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images," *Signal, Image and Video Processing*, vol. 9, no. 2, pp. 409–425, 2015.

[136] A. M. Hadi, K. T. Mouchaers, I. Schalij, K. Grunberg, G. A. Meijer, A. Vonk-Noordegraaf, W. J. van der Laarse, and J. A. Beliën, "Rapid quantification of myocardial fibrosis: a new macro-based automated analysis," *Cellular Oncology*, vol. 34, no. 4, pp. 343–354, 2011.

[137] R. Marée, S. Dallongeville, J.-C. Olivo-Marin, and V. Meas-Yedid, "An approach for detection of glomeruli in multisite digital pathology," in *IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016*. IEEE, 2016, pp. 1033–1036.

[138] I. Pitas and N. Nikolaidis, *3-D image processing algorithms*. Wiley, 2000.

[139] P. J. Baldevbhai and R. Anand, "Color image segmentation for medical images using l* a* b* color space," *IOSR Journal of Electronics and Communication Engineering*, vol. 1, no. 2, pp. 24–45, 2012.

[140] M. Tkalcic and J. F. Tasic, *Colour spaces: perceptual, historical and applicational background*. IEEE, 2003, vol. 1.

[141] M. V. Daithankar, K. J. Karande, and A. D. Harale, "Analysis of skin color models for face detection," in *International Conference on Communications and Signal Processing (ICCSP), 2014*. IEEE, 2014, pp. 533–537.

[142] J. Dong, J. Li, A. Fu, and H. Lv, "Automatic segmentation for ovarian cancer immunohistochemical image based on yuv color space," in *International Conference on Biomedical Engineering and Computer Science (ICBECS), 2010*. IEEE, 2010, pp. 1–4.

[143] K. Z. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1153–1163, 2006.

[144] A. Sadr, M. Jahed, P. Salehian, and A. Eslami, "Leukocyte's nucleus segmentation using active contour in ycbcr colour space," in *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2010*. IEEE, 2010, pp. 257–260.

[145] B. Liu, W. Wang, and X. Fang, "Automatic extraction of positive cells in tumor immunohistochemical pathology image based on ycbcr," in *The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006*, vol. 2. IEEE, 2006, pp. 9708–9712.

[146] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333–343, 2011.

[147] A. A. Nasir, M. Mashor, and H. Rosline, "Unsupervised colour segmentation of white blood cell for acute leukaemia images," in *IEEE International Conference on Imaging Systems and Techniques (IST), 2011*. IEEE, 2011, pp. 142–145.

[148] A. A. Salihah, M. Mashor, N. H. Harun, A. A. Abdullah, and H. Rosline, "Improving colour image segmentation on acute myelogenous leukaemia images using contrast enhancement techniques," in *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2010*. IEEE, 2010, pp. 246–251.

[149] J. Wu, P. Zeng, Y. Zhou, and C. Olivier, "A novel color image segmentation method and its application to white blood cell image analysis," in *8th International Conference on Signal Processing, 2006*, vol. 2. IEEE, 2006.

[150] J. Duan and L. Yu, "A wbc segmentation method based on hsi color space," in *4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), 2011*. IEEE, 2011, pp. 629–632.

[151] L.-W. Hao, W.-X. Hong, and C.-L. Hu, "A novel auto-segmentation scheme for colored leukocyte images," in *First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA), 2010*. IEEE, 2010, pp. 916–919.

[152] H. N. Hazlyna, M. Y. Mashor, N. R. Mokhtar, A. A. Salihah, R. Hassan, R. A. A. Raof, and M. K. Osman, "Comparison of acute leukemia image segmentation using hsi and rgb color space," in *10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), 2010*. IEEE, 2010, pp. 749–752.

[153] R. A. A. Raof, M. Y. Mashor, R. B. Ahmad, S. S. M. Noor, and M. K. Osman, "Comparison of colour thresholding method using rgb and hsi information for ziehl-neelsen sputum slide images," in *10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), 2010*. IEEE, 2010, pp. 724–727.

[154] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, "Med-node: a computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6578–6585, 2015.

[155] A. P. Dhawan and A. Sim, "Segmentation of images of skin lesions using color and texture information of surface pigmentation," *Computerized Medical Imaging and Graphics*, vol. 16, no. 3, pp. 163–177, 1992.

[156] X. Wang, B. Zhang, Z. Yang, H. Wang, and D. Zhang, "Statistical analysis of tongue images for feature extraction and diagnostics," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5336–5347, 2013.

[157] H. Dai and X. Li, "The color components' computing on different color spaces for image segmentation of hematocyte," in *International Symposium on Intelligence Information Processing and Trusted Computing (IPTC), 2010*. IEEE, 2010, pp. 75–78.

[158] M. Sotaquira, L. Rueda, and R. Narvaez, "Detection and quantification of bacilli and clusters present in sputum smear samples: a novel algorithm for pulmonary tuberculosis diagnosis," in *International Conference on Digital Image Processing, 2009*. IEEE, 2009, pp. 117–121.

[159] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and

cell splitting," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1661–1677, 2011.

[160] X. Wang, B. Zhang, Z. Guo, and D. Zhang, "Facial image medical analysis system using quantitative chromatic feature," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3738–3746, 2013.

[161] V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, "Computer-aided diagnosis of psoriasis skin images with hos, texture and color features: a first comparative study of its kind," *Computer Methods and Programs in Biomedicine*, vol. 126, pp. 98–109, 2016.

[162] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "An automatic color segmentation algorithm with application to identification of skin tumor borders," *Computerized Medical Imaging and Graphics*, vol. 16, no. 3, pp. 227–235, 1992.

[163] C. H. Li and P. C. Yuen, "Tongue image matching using color content," *Pattern Recognition*, vol. 35, no. 2, pp. 407–419, 2002.

[164] A. P. Dhawan, Y. Chitre, and C. Kaiser-Bonasso, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 246–259, 1996.

[165] N. Ramesh, J.-H. Yoo, and I. Sethi, "Thresholding based on histogram approximation," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 142, no. 5, pp. 271–279, 1995.

[166] Q. Wu, F. Merchant, and K. Castleman, *Microscope image processing*. Elsevier, 2010.

[167] R. S. Kumar, A. Verma, and J. Singh, "Color image segmentation and multi-level thresholding by maximization of conditional entrophy," *International Journal of Signal Processing*, vol. 3, no. 1, pp. 121–125, 2006.

[168] M. Sezgin and R. Taşaltın, "A new dichotomization technique to multilevel thresholding devoted to inspection applications," *Pattern Recognition Letters*, vol. 21, no. 2, pp. 151–161, 2000.

[169] W. Boecker, W.-U. Muller, and C. Streffer, "Comparison of different automatic threshold algorithms for image segmentation in microscope images," in *Applications of Digital Image Processing XVIII*, vol. 2564. International Society for Optics and Photonics, 1995, pp. 230–242.

[170] Y. Zhang, H. Qu, and Y. Wang, "Adaptive image segmentation based on fast thresholding and image merging," in *16th International Conference on Artificial Reality and Telexistence–Workshops, 2006. ICAT'06*. IEEE, 2006, pp. 308–311.

[171] S. U. Lee, S. Y. Chung, and R. H. Park, "A comparative performance study of several global thresholding techniques for segmentation," *Computer Vision, Graphics, and Image Processing*, vol. 52, no. 2, pp. 171–190, 1990.

[172] B. J. Schachter, L. S. Davis, and A. Rosenfeld, "Some experiments in image segmentation by clustering of local feature values," *Pattern Recognition*, vol. 11, no. 1, pp. 19–28, 1979.

[173] W.-X. Kang, Q.-Q. Yang, and R.-P. Liang, "The comparative research on image segmentation algorithms," in *First International Workshop on Education Technology and Computer Science, 2009. ETCS'09*, vol. 2. IEEE, 2009, pp. 703–707.

[174] X. M. Pardo and D. Cabello, "Biomedical active segmentation guided by edge saliency," *Pattern Recognition Letters*, vol. 21, no. 6-7, pp. 559–572, 2000.

[175] M. Rastgarpour and J. Shanbehzadeh, "Application of ai techniques in medical image segmentation and novel categorization of available methods and," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2011*. Citeseer, 2011.

[176] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 3rd ed. Pearson, 2007.

[177] G. Castellano, L. Bonilha, L. Li, and F. Cendes, "Texture analysis of medical images," *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.

[178] J. Rogowska, "Overview and fundamentals of medical image segmentation," *Handbook of Medical Imaging, Processing and Analysis*, pp. 69–85, 2000.

[179] H.-D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[180] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 260–280, 2008.

[181] H. T. Nguyen, M. Worring, and R. Van Den Boomgaard, "Watersnakes: Energy-driven watershed segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 330–342, 2003.

[182] A. Elnakib, G. Gimel'farb, J. S. Suri, and A. El-Baz, "Medical image segmentation: a brief survey," in *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*. Springer, 2011, pp. 1–39.

[183] A. Kassner and R. Thornhill, "Texture analysis: a review of neurologic mr imaging applications," *American Journal of Neuroradiology*, vol. 31, no. 5, pp. 809–816, 2010.

[184] J. F. Silverman and D. B. Cooper, "Bayesian clustering for unsupervised estimation of surface and texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 482–495, 1988.

[185] R. M. Haralick, K. Shanmugam, I. Dinstein *et al.*, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[186] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[187] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 1, pp. 55–73, 1990.

[188] D. Blostein and N. Ahuja, "Shape from texture: Integrating texture-element extraction and surface estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1233–1251, 1989.

[189] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[190] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[191] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005*, vol. 1. IEEE, 2005, pp. 886–893.

[192] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[193] P. Maji, M. K. Kundu, and B. Chanda, "Second order fuzzy measure and weighted co-occurrence matrix for segmentation of brain mr images," *Fundamenta Informaticae*, vol. 88, no. 1-2, pp. 161–176, 2008.

[194] A. M. Khuzi, R. Besar, W. W. Zaki, and N. Ahmad, "Identification of masses in digital mammogram using gray level co-occurrence matrices," *Biomedical Imaging and Intervention Journal*, vol. 5, no. 3, 2009.

[195] H. Qureshi, N. M. Rajpoot, T. Nattkemper, and V. Hans, "A robust adaptive wavelet-based method for classification of meningioma histology images," 2009.

[196] A. Katouzian, B. Baseri, E. E. Konofagou, and A. F. Laine, "Automatic detection of blood versus non-blood regions on intravascular ultrasound (ivus) images using wavelet packet signatures," in *Medical Imaging 2008: Ultrasonic Imaging and Signal Processing*, vol. 6920. International Society for Optics and Photonics, 2008, p. 69200H.

[197] A. Montillo, D. Metaxas, and L. Axel, "Extracting tissue deformation using gabor filter banks," in *Medical Imaging 2004: Physiology, Function, and Structure from*

Medical Images, vol. 5369. International Society for Optics and Photonics, 2004, pp. 1–10.

[198] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli, "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 4, pp. 344–357, 2003.

[199] R. S. Alomari, S. Ghosh, V. Chaudhary, and O. Al-Kadi, "Local binary patterns for stromal area removal in histology images," in *Medical Imaging 2012: Computer-Aided Diagnosis*, vol. 8315. International Society for Optics and Photonics, 2012, p. 831524.

[200] K. Masood and N. Rajpoot, "Texture based classification of hyperspectral colon biopsy samples using clbp," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009. ISBI'09.* IEEE, 2009, pp. 1011–1014.

[201] O. S. Al-Kadi, "Texture measures combination for improved meningioma classification of histopathological images," *Pattern Recognition*, vol. 43, no. 6, pp. 2043–2053, 2010.

[202] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Conference on Artificial Intelligence in Medicine in Europe.* Springer, 2009, pp. 126–135.

[203] J. C. Caicedo, F. A. González, and E. Romero, "Content-based histopathology image retrieval using a kernel-based semantic annotation framework," *Journal of Biomedical Informatics*, vol. 44, no. 4, pp. 519–528, 2011.

[204] G. Díaz and E. Romero, "Micro-structural tissue analysis for automatic histopathological image annotation," *Microscopy Research and Technique*, vol. 75, no. 3, pp. 343–358, 2012.

[205] S. Doyle, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "A boosting cascade for automated detection of prostate cancer from digitized histology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, 2006, pp. 504–511.

[206] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008.* IEEE, 2008, pp. 496–499.

[207] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinformatics*, vol. 13, no. 1, p. 282, 2012.

[208] H. Kalkan, M. Nap, R. P. Duin, and M. Loog, "Automated classification of local patches in colon histopathology," in *21st International Conference on Pattern Recognition (ICPR), 2012.* IEEE, 2012, pp. 61–64.

[209] W. Wang, J. A. Ozolek, and G. K. Rohde, "Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images," *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry*, vol. 77, no. 5, pp. 485–494, 2010.

[210] S. Salman, Z. Ma, S. Mohanty, S. Bhele, Y.-T. Chu, B. Knudsen, and A. Gertych, "A machine learning approach to identify prostate cancer areas in complex histological images," in *Information Technologies in Biomedicine, Volume 3*. Springer, 2014, pp. 295–306.

[211] Y. Song, W. Cai, S. Huh, M. Chen, T. Kanade, Y. Zhou, and D. Feng, "Discriminative data transform for image feature extraction and classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 452–459.

[212] Y. Xu, L. Jiao, S. Wang, J. Wei, Y. Fan, M. Lai, and E. I.-c. Chang, "Multi-label classification for colon cancer using histopathological images," *Microscopy Research and Technique*, vol. 76, no. 12, pp. 1266–1277, 2013.

[213] T. Wan, J. Cao, J. Chen, and Z. Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomputing*, vol. 229, pp. 34–44, 2017.

[214] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. Tizhoosh, "A comparative study of cnn, bovw and lbp for classification of histopathological images," *arXiv preprint arXiv:1710.01249*, 2017.

[215] M. M. R. Krishnan, V. Venkatraghavan, U. R. Acharya, M. Pal, R. R. Paul, L. C. Min, A. K. Ray, J. Chatterjee, and C. Chakraborty, "Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm," *Micron*, vol. 43, no. 2-3, pp. 352–364, 2012.

[216] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. T. Brunyé, and J. G. Elmore, "Localization of diagnostically relevant regions of interest in whole slide images: A comparative study," *Journal of Digital Imaging*, vol. 29, no. 4, pp. 496–506, 2016.

[217] C. M. Lopez and S. Agaian, "A new set of wavelet-and fractals-based features for gleason grading of prostate cancer histopathology images," in *Image Processing: Algorithms and Systems XI*, vol. 8655. International Society for Optics and Photonics, 2013, p. 865516.

[218] M. M. R. Krishnan, P. Shah, A. Choudhary, C. Chakraborty, R. R. Paul, and A. K. Ray, "Textural characterization of histopathological images for oral sub-mucous fibrosis detection," *Tissue and Cell*, vol. 43, no. 5, pp. 318–330, 2011.

[219] M. M. R. Krishnan, C. Chakraborty, and A. K. Ray, "Wavelet based texture classification of oral histopathological sections," *International Journal of Microscopy, Science, Technology, Applications and Education*, vol. 2, no. 4, pp. 897–906, 2010.

[220] B. Ramamurthy and K. Chandran, "Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms," *Journal of Computer Science*, vol. 8, no. 7, p. 1070, 2012.

[221] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 496–506, 2015.

[222] A. Cruz-Roa, J. C. Caicedo, and F. A. González, "Visual pattern mining in histology image collections using bag of features," *Artificial Intelligence in Medicine*, vol. 52, no. 2, pp. 91–106, 2011.

[223] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical Image Analysis*, vol. 14, no. 2, pp. 227–241, 2010.

[224] X. Zhang, H. Su, L. Yang, and S. Zhang, "Weighted hashing with multiple cues for cell-level analysis of histopathological images," in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 303–314.

[225] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial Intelligence in Medicine*, vol. 49, no. 2, pp. 117–125, 2010.

[226] M. Moradi, P. Abolmaesoumi, and P. Mousavi, "Deformable registration using scale space keypoints," in *Medical Imaging 2006: Image Processing*, vol. 6144. International Society for Optics and Photonics, 2006, p. 61442G.

[227] Y. Srinivasan, D. Hernes, B. Tulpule, S. Yang, J. Guo, S. Mitra, S. Yagneswaran, B. Nutter, J. Jeronimo, B. Phillips *et al.*, "A probabilistic approach to segmentation and classification of neoplasia in uterine cervix images using color and geometric features," in *Medical Imaging 2005: Image Processing*, vol. 5747. International Society for Optics and Photonics, 2005, pp. 995–1004.

[228] A. Nedzved, A. Belotserkovsky, T. Lehmann, and S. Ablameyko, "Morphometrical feature extraction on color histological images for oncological diagnostics," in *5th International Conference on Biomedical Engineering*, 2007, pp. 14–16.

[229] J. Angulo, J. Klossa, and G. Flandrin, "Ontology-based lymphocyte population description using mathematical morphology on colour blood images," *Cellular and Molecular Biology*, vol. 52, no. 6, pp. 2–15, 2006.

[230] L. Cooper, O. Sertel, J. Kong, G. Lozanski, K. Huang, and M. Gurcan, "Feature-based registration of histopathology images with different stains: An application for computerized follicular lymphoma prognosis," *Computer Methods and Programs in Biomedicine*, vol. 96, no. 3, pp. 182–192, 2009.

[231] A. Nedzved, S. Ablameyko, and I. Pitas, "Morphological segmentation of histology cell images," in *15th International Conference on Pattern Recognition Proceedings, 2000*, vol. 1. IEEE, 2000, pp. 500–503.

[232] J. T. Kwak, S. M. Hewitt, S. Sinha, and R. Bhargava, "Multimodal microscopy for automated histologic analysis of prostate cancer," *BMC Cancer*, vol. 11, no. 1, p. 62, 2011.

[233] H. Sharma, N. Zerbe, S. Lohmann, K. Kayser, O. Hellwich, and P. Hufnagl, "A review of graph-based methods for image analysis in digital histopathology," *Diagnostic Pathology*, 2015.

[234] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008.* IEEE, 2008, pp. 284–287.

[235] J. Naik, S. Doyle, A. Basavanhally, S. Ganesan, M. D. Feldman, J. E. Tomaszewski, and A. Madabhushi, "A boosted distance metric: application to content based image retrieval and classification of digitized histopathology," in *Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260. International Society for Optics and Photonics, 2009, p. 72603F.

[236] A. Basavanhally, S. Ganesan, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, and A. Madabhushi, "Multi-field-of-view framework for distinguishing tumor grade in er+ breast cancer from entire histopathology slides," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2089–2099, 2013.

[237] C. M. Lopez, S. Agaian, I. Sanchez, A. Almuntashri, O. Zinalabdin, A. Rikabi, and I. Thompson, "Exploration of efficacy of gland morphology and architectural features in prostate cancer gleason grading," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2012.* IEEE, 2012, pp. 2849–2854.

[238] M. Guillaud, D. Cox, K. Adler-Storthz, A. Malpica, G. Staerkel, J. Matisic, D. Van Niekerk, N. Poulin, M. Follen, and C. MacAulay, "Exploratory analysis of quantitative histopathology of cervical intraepithelial neoplasia: Objectivity, reproducibility, malignancy-associated changes, and human papillomavirus," *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, vol. 60, no. 1, pp. 81–89, 2004.

[239] P. A. Lachenbruch and M. Goldstein, "Discriminant analysis," *Biometrics*, pp. 69–85, 1979.

[240] J. He, H.-J. Hu, R. Harrison, P. C. Tai, and Y. Pan, "Transmembrane segments prediction and understanding using support vector machine and decision tree," *Expert Systems with Applications*, vol. 30, no. 1, pp. 64–72, 2006.

[241] H. K. Sok, M. P.-L. Ooi, and Y. C. Kuang, "Sparse alternating decision tree," *Pattern Recognition Letters*, vol. 60, pp. 57–64, 2015.

[242] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[243] H. K. Sok, M. P.-L. Ooi, Y. C. Kuang, and S. Demidenko, "Multivariate alternating decision trees," *Pattern Recognition*, vol. 50, pp. 195–209, 2016.

[244] T. M. Mitchell *et al.*, "Machine learning. wcb," 1997.

[245] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542–1550, 2002.

[246] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, and H. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition*, vol. 39, no. 4, pp. 646–668, 2006.

[247] Y.-J. Zhang, "An overview of image and video segmentation in the last 40 years," in *Advances in Image and Video Segmentation*. IGI Global, 2006, pp. 1–16.

[248] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, Jan 2000.

[249] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of Medical Physics/Association of Medical Physicists of India*, vol. 35, no. 1, p. 3, 2010.

[250] B. R. Kumar, D. K. Joseph, and T. Sreenivas, "Teager energy based blood cell segmentation," in *14th International Conference on Digital Signal Processing, 2002*, vol. 2. IEEE, 2002, pp. 619–622.

[251] T. J. Hebert, "Fast iterative segmentation of high resolution medical images," *IEEE Transactions on Nuclear Science*, vol. 44, no. 3, pp. 1362–1367, 1997.

[252] J. C. Bezdek, L. Hall, and L. Clarke, "Review of mr image segmentation techniques using pattern recognition," *Medical Physics*, vol. 20, no. 4, pp. 1033–1048, 1993.

[253] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[254] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.

[255] L. D. Cohen, "On active contour models and balloons," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 211–218, 1991.

[256] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.

[257] S. Z. Li, "Markov random field models in computer vision," in *European Conference on Computer Vision*. Springer, 1994, pp. 361–370.

[258] J. Liu and Y.-H. Yang, "Multiresolution color image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 7, pp. 689–700, 1994.

[259] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 259–302, 1986.

[260] A. El-Baz and A. A. Farag, "Image segmentation using gmrf models: parameters estimation and applications," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2. IEEE, 2003, pp. II–177.

[261] S. Haykin and N. Network, "A comprehensive foundation," *Neural Networks*, vol. 2, no. 2004, p. 41, 2004.

[262] L. O. Hall, A. M. Bensaid, L. P. Clarke, R. P. Velthuizen, M. S. Silbiger, and J. C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 672–682, 1992.

[263] C. Vijayakumar, G. Damayanti, R. Pant, and C. Sreedhar, "Segmentation and grading of brain tumors on apparent diffusion coefficient images using self-organizing maps," *Computerized Medical Imaging and Graphics*, vol. 31, no. 7, pp. 473–484, 2007.

[264] S. Chaplot, L. Patnaik, and N. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network," *Biomedical Signal Processing and Control*, vol. 1, no. 1, pp. 86–92, 2006.

[265] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, "Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks," in *European Conference on Computer Vision*. Springer, 2008, pp. 69–82.

[266] K. Li, C. Zou, S. Bu, Y. Liang, J. Zhang, and M. Gong, "Multi-modal feature fusion for geographic image annotation," *Pattern Recognition*, vol. 73, pp. 1–14, 2018.

[267] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[268] W. K. Tey, Y. C. Kuang, M. P.-L. Ooi, and J. J. Khoo, "Automated quantification of renal interstitial fibrosis for computer-aided diagnosis: A comprehensive tissue structure segmentation method," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 109–120, 2018.

[269] H. Yamada, C. Merritt, and T. Kasvand, "Recognition of kidney glomerulus by dynamic programming matching method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 731–737, 1988.

[270] Y. Zhao, E. F. Black, L. Marini, K. McHenry, N. Kenyon, R. Patil, A. Balla, and A. Bartholomew, "Automatic glomerulus extraction in whole slide images towards computer aided diagnosis," in *IEEE 12th International Conference on e-Science (e-Science), 2016*. IEEE, 2016, pp. 165–174.

[271] T. Kotyk, N. Dey, A. S. Ashour, D. Balas-Timar, S. Chakraborty, A. S. Ashour, and J. M. R. Tavares, "Measurement of glomerulus diameter and bowman's space width of renal albino rats," *Computer Methods and Programs in Biomedicine*, vol. 126, pp. 143–153, 2016.

[272] J. Zhang and J. Fan, "Glomerulus extraction based on genetic algorithm and watershed transform," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006*. IEEE, 2006, pp. 4863–4866.

[273] J. Zhang, H. Zhu, X. Qian, and T. Huang, "Genetic algorithm for edge extraction of glomerulus area," in *International Conference on Information Acquisition Proceedings, 2004*. IEEE, 2004, pp. 335–338.

[274] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 6, pp. 583–598, 1991.

[275] M. Xie and Z. Ma, "Research of edge detection for b-spline wavelet," *Acta Electronica Sinica*, vol. 27, pp. 106–108, 1999.

[276] J. Zhang and Q. Zhang, "Color image segmentation based on wavelet transformation and s ofm neural network," in *IEEE International Conference on Robotics and Biomimetics, 2007*. IEEE, 2007, pp. 1778–1781.

[277] A. Pedraza, J. Gallego, S. Lopez, L. Gonzalez, A. Laurinavicius, and G. Bueno, "Glomerulus classification with convolutional neural networks," in *Annual Conference on Medical Image Understanding and Analysis*. Springer, 2017, pp. 839–849.

[278] I. Kamenetsky, R. M. Rangayyan, and H. Benediktsson, "Segmentation and analysis of the glomerular basement membrane using the split and merge method," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008*. IEEE, 2008, pp. 3064–3067.

[279] B. Ginley, J. E. Tomaszewski, R. Yacoub, F. Chen, and P. Sarder, "Unsupervised labeling of glomerular boundaries using gabor filters and statistical testing in renal histology," *Journal of Medical Imaging*, vol. 4, no. 2, p. 021102, 2017.

[280] Y. Hirohashi, R. Relator, T. Kakimoto, R. Saito, Y. Horai, A. Fukunari, H. Utsumi, K. Okada, and T. Kato, "Automated quantitative image analysis of glomerular desmin immunostaining as a sensitive injury marker in spontaneously diabetic torii rats," *Journal of Biomedical Image Processing*, vol. 1, no. 1, pp. 20–8, 2014.

[281] T. Kato, R. Relator, H. Ngouv, Y. Hirohashi, O. Takaki, T. Kakimoto, and K. Okada, "Segmental hog: new descriptor for glomerulus detection in kidney microscopy image," *BMC Bioinformatics*, vol. 16, no. 1, p. 316, 2015.

[282] M. Gadermayr, B. M. Klinkhammer, P. Boor, and D. Merhof, "Do we need large annotated training data for detection applications in biomedical imaging? a case study in renal glomeruli detection," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2016, pp. 18–26.

[283] D. Daunoravicius, J. Besusparis, E. Zurauskas, A. Laurinaviciene, D. Bironaite, S. Pankuweit, B. Plancoulaine, P. Herlin, J. Bogomolovas, V. Grabauskiene *et al.*, "Quantification of myocardial fibrosis by digital image analysis and interactive stereology," *Diagnostic Pathology*, vol. 9, no. 1, p. 114, 2014.

[284] K. Marten, V. Dicken, C. Kneitz, M. Hoehmann, W. Kenn, D. Hahn, and C. Engelke, "Computer-assisted quantification of interstitial lung disease associated with rheumatoid arthritis: preliminary technical validation," *European Journal of Radiology*, vol. 72, no. 2, pp. 278–283, 2009.

[285] M. D. Taylor, J. R. Roberts, A. F. Hubbs, M. J. Reasor, and J. M. Antonini, "Quantitative image analysis of drug-induced lung fibrosis using laser scanning confocal microscopy," *Toxicological Sciences*, vol. 67, no. 2, pp. 295–302, 2002.

[286] M. Masseroli, T. Caballero, F. O'valle, R. M. Del Moral, A. Pérez-Milena, and R. G. Del Moral, "Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis," *Journal of Hepatology*, vol. 32, no. 3, pp. 453–464, 2000.

[287] M. Masseroli, F. O'Valle, M. Andújar, C. Ramírez, M. Gómez-Morales, M. Aguilar, D. Aguilar, M. Rodriguez-Puyol, R. M. Del *et al.*, "Design and validation of a new image analysis method for automatic quantification of interstitial fibrosis and glomerular morphometry." *Laboratory Investigation; A Journal of Technical Methods and Pathology*, vol. 78, no. 5, pp. 511–522, 1998.

[288] L. B. Sant'Anna, N. Sant'Anna, and O. Parolini, "Application of computer-assisted image analysis for identifying and quantifying liver fibrosis in an experimental model," *Journal of Computational Interdisciplinary Sciences*, vol. 2, no. 2, pp. 139–48, 2011.

[289] K. Meejaroen, C. Chaweechan, W. Khodsiri, V. Khu-smith, U. Watchareeruetai, P. Sornmagura, and T. Kittiyakara, "Detection of fibrosis in liver biopsy images by using bayesian classifier," in *7th International Conference on Knowledge and Smart Technology (KST), 2015*. IEEE, 2015, pp. 184–189.

[290] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition*, vol. 25, no. 10, pp. 1231–1240, 1992.

[291] A. Servais, V. Meas-Yedid, M. Buchler, E. Morelon, J.-C. Olivo-Marin, Y. Lebranchu, C. Legendre, and E. Thervet, "Quantification of interstitial fibrosis by image analysis on routine renal biopsy in patients receiving cyclosporine," *Transplantation*, vol. 84, no. 12, pp. 1595–1601, 2007.

[292] F. Moreso, D. Serón, J. Vitriá, J. M. Grinyó, F. M. Colomé-Serra, N. Parés, and J. Serra, "Quantification of interstitial chronic renal damage by means of texture analysis," *Kidney International*, vol. 46, no. 6, pp. 1721–1727, 1994.

[293] M. Klapczynski, G. D. Gagne, S. J. Morgan, K. J. Larson, B. E. LeRoy, E. A. Blomme, B. F. Cox, and E. W. Shek, "Computer-assisted imaging algorithms facilitate histomorphometric quantification of kidney damage in rodent renal failure models," *Journal of Pathology Informatics*, vol. 3, 2012.

[294] J. M. Gauch and C. W. Hsia, "Comparison of three-color image segmentation algorithms in four color spaces," in *Visual Communications and Image Processing'92*, vol. 1818. International Society for Optics and Photonics, 1992, pp. 1168–1182.

[295] R. I.-R. BT, "Studio encoding parameters of digital television for standard 4: 3 and wide-screen 16: 9 aspect ratios," 1990.

[296] C. CIE, "Commission internationale de l'eclairage proceedings, 1931," *Cambridge University Press Cambridge*, 1932.

[297] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[298] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[299] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[300] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[301] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338–1350, 2001.

[302] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[303] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina, "An efficient algorithm for fractal analysis of textures," in *25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2012*. IEEE, 2012, pp. 39–46.

[304] X. Wei, "Gray level run length matrix toolbox v1. 0," *Software, Beijing Aeronautical Technology Research Center*, 2007.

[305] N. Aggarwal and R. Agrawal, "First and second order statistics features for classification of magnetic resonance brain images," *Journal of Signal and Information Processing*, vol. 3, no. 02, p. 146, 2012.

[306] I. Daubechies, *Ten lectures on wavelets*. Siam, 1992, vol. 61.

[307] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.

[308] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability." *Psychological Bulletin*, vol. 86, no. 2, p. 420, 1979.

[309] K. A. Hallgren, "Computing inter-rater reliability for observational data: an overview and tutorial," *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, p. 23, 2012.

[310] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

[311] A. Farris, S. Chan, J. Climenhaga, B. Adam, C. Bellamy, D. Seron, R. Colvin, J. Reeve, and M. Mengel, "Banff fibrosis study: multicenter visual assessment and computerized analysis of interstitial fibrosis in kidney biopsies," *American Journal of Transplantation*, vol. 14, no. 4, pp. 897–907, 2014.

[312] K. Solez, R. Colvin, L. Racusen, M. Haas, B. Sis, M. Mengel, P. Halloran, W. Baldwin, G. Banfi, A. Collins *et al.*, "Banff 07 classification of renal allograft pathology: updates and future directions," *American Journal of Transplantation*, vol. 8, no. 4, pp. 753–760, 2008.

[313] P. Kubelka, "New contributions to the optics of intensely light-scattering materials. part i," *Josa*, vol. 38, no. 5, pp. 448–457, 1948.

[314] R. A. Fisher, *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 2006.

[315] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients." *Psychological Methods*, vol. 1, no. 1, p. 30, 1996.

[316] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[317] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological Bulletin*, vol. 76, no. 5, p. 378, 1971.