Computer Methods and Programs in Biomedicine 155 (2018) 109–120

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Automated quantification of renal interstitial fibrosis for computer-aided diagnosis: A comprehensive tissue structure segmentation method

Wei Keat Tey[a,*], Ye Chow Kuang[a], Melanie Po-Leen Ooi[b], Joon Joon Khoo[c]

[a] Advanced Engineering Platform and Department of Electrical and Computer Systems Engineering, School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway 47500, Malaysia
[b] Unitec Institute of Technology, 139 Carrington Road, Mount Albert, Auckland 1025, New Zealand
[c] Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Bandar Sunway 47500, Malaysia

ARTICLE INFO

Article history:
Received 28 July 2017
Revised 1 December 2017
Accepted 11 December 2017

Keywords:
Renal interstitial fibrosis
Medical image analysis
Computer aided diagnosis
Tissue structure segmentation

ABSTRACT

Interstitial fibrosis in renal biopsy samples is a scarring tissue structure that may be visually quantified by pathologists as an indicator to the presence and extent of chronic kidney disease. The standard method of quantification by visual evaluation presents reproducibility issues in the diagnoses. This study proposes an automated quantification system for measuring the amount of interstitial fibrosis in renal biopsy images as a consistent basis of comparison among pathologists. The system extracts and segments the renal tissue structures based on colour information and structural assumptions of the tissue structures. The regions in the biopsy representing the interstitial fibrosis are deduced through the elimination of non-interstitial fibrosis structures from the biopsy area and quantified as a percentage of the total area of the biopsy sample. A ground truth image dataset has been manually prepared by consulting an experienced pathologist for the validation of the segmentation algorithms. The results from experiments involving experienced pathologists have demonstrated a good correlation in quantification result between the automated system and the pathologists' visual evaluation. Experiments investigating the variability in pathologists also proved the automated quantification error rate to be on par with the average intra-observer variability in pathologists' quantification.

Background and Objective: Interstitial fibrosis in renal biopsy samples is a scarring tissue structure that may be visually quantified by pathologists as an indicator to the presence and extent of chronic kidney disease. The standard method of quantification by visual evaluation presents reproducibility issues in the diagnoses due to the uncertainties in human judgement.

Methods: An automated quantification system for accurately measuring the amount of interstitial fibrosis in renal biopsy images is presented as a consistent basis of comparison among pathologists. The system identifies the renal tissue structures through knowledge-based rules employing colour space transformations and structural features extraction from the images. In particular, the renal glomerulus identification is based on a multiscale textural feature analysis and a support vector machine. The regions in the biopsy representing interstitial fibrosis are deduced through the elimination of non-interstitial fibrosis structures from the biopsy area. The experiments conducted evaluate the system in terms of quantification accuracy, intra- and inter-observer variability in visual quantification by pathologists, and the effect introduced by the automated quantification system on the pathologists' diagnosis.

Results: A 40-image ground truth dataset has been manually prepared by consulting an experienced pathologist for the validation of the segmentation algorithms. The results from experiments involving experienced pathologists have demonstrated an average error of 9 percentage points in quantification result between the automated system and the pathologists' visual evaluation. Experiments investigating the variability in pathologists involving samples from 70 kidney patients also proved the automated quantification error rate to be on par with the average intra-observer variability in pathologists' quantification.

Conclusions: The accuracy of the proposed quantification system has been validated with the ground truth dataset and compared against the pathologists' quantification results. It has been shown that the

* Corresponding author.
E-mail address: tey.wei.keat@monash.edu (W.K. Tey).

https://doi.org/10.1016/j.cmpb.2017.12.004

correlation between different pathologists' estimation of interstitial fibrosis area has significantly improved, demonstrating the effectiveness of the quantification system as a diagnostic aide.

1. Introduction

The annual death rate of chronic kidney disease has seen a sharp rise from 15.7 to 16.3 per 100,000 between the years 1990 and 2010. According to the 2010 Global Burden of Diseases study, this rate of increase was only second to that of HIV and AIDS [1]. Unfortunately, this trend is expected to increase with the growth of the global elderly population, especially in developing countries.

The diagnosis of kidney disease is typically carried out by blood and urine tests, while biopsy samples are required in certain cases when the specific cause of the diseases is unclear or abnormal [2]. These biopsy samples will be stained and observed for the presence of any pathological structures, and one of the main identifying structures of chronic kidney disease is interstitial fibrosis. The visual evaluation of the biopsy sample is conducted by pathologists under a light microscope in a semi-quantitative manner, in which a score of the approximate amount of interstitial fibrosis observed is produced. One of the only standard for this semi-quantitative interstitial fibrosis quantification is the Banff schema commonly used for the renal allograft rejection grades classification [3]. Despite the existence of this largely generalised scoring standard, visual scoring is still very much subject to the pathologists' intra- and inter-observer variability, and it may not be repeatable or reproducible [4–6]. This variability arises from the fact that even experts tend to produce an overestimation or underestimation in ambiguous cases [7]. These middle areas (or 'grey area') of which the decision threshold is based heavily upon the judgement of each observer are the main research issue in this study.

Recent advances in computerised tools in the biomedical area have spurred the development of automated image analysis systems to aid pathologists in accurately quantifying interstitial fibrosis. The precision and meticulousness of the computer system provide an advantage over human judgment in medical image analysis, especially in the quantification of tissue structure area in digital images. Specifically, in the case of this research, the colour information from chemically-stained biopsy slides serves as the main input to the computer system for processing. In computer vision applications, a colour image can be represented in several different colour spaces [8]. A colour space is effectively the description of the range of colours that can be reproduced in an image, described by a mathematical relationship that correlates colours to physical ink colours or illuminated displays on digital display devices. Various colour spaces present different representation of colour information that suit specific applications. Colour space transformation finds its application in the analysis of stained histopathological slides for the medical and healthcare area [9–15].

The automatic quantification of fibrosis has been well explored for applications in the liver [16–18], lung [19,20], and heart [21,22]. Research on computer-aided quantification of interstitial fibrosis in the kidney has also been carried out, with the majority based on the premise of point counting. In other words, the total pixel area corresponding to fibrosis is calculated as a percentage of the whole biopsy tissue sample. For example, Masseroli et al. [23] 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. [24] to produce a binary image for subsequent classification of features based on their shape and size. Similarly, Servais et al. [25] also presented an automatic quantification method that extracts the green-coloured region of trichrome-stained renal biopsies, which is the colour indicative of interstitial fibrosis for the stain. On the other hand, Moreso et al. [26] employed a texture analysis method instead of attempting to segment tubulointerstitial structure 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 interstitial fibrosis quantification cannot be achieved without positive identification of the segmented structures or regions of interest. This approach has been adopted by Klapczynski et al. [27] in rat renal tissue specimens, while a more comprehensive system for identifying human kidney structures was presented by Meas-Yedid et al. [28]. 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 $I_1H_2H_3$ colour space. Since the interstitial fibrosis structure is not the sole structure stained green by the trichrome stain, the system proceeded to remove other non-interstitial fibrosis structures such as the renal capsule, basement membrane, glomeruli, and blood vessels. However, segmentation by 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. Aside from that, 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 amount of interstitial fibrosis, especially tubulointerstitial fibrosis, has been shown to be a good predictor of progression of end-stage renal disease [29]. When there is a damage or injury to the kidney, the tissue responds by forming fibrosis in both the glomeruli and interstitium. However, it is not as useful to quantitate glomerulosclerosis because the glomerulosclerosis will be replaced by interstitial fibrosis in later stages of the disease. Moreover, previous studies have shown that a decrease in glomerular filtration rate is better correlated with tubulointerstitial injury than with glomerulosclerosis [30,31]. Thus, having an accurate quantification of the amount of interstitial fibrosis will help in predicting the time lapse before renal failure sets in for a patient. To the best of our knowledge, there has not been an automated quantification system for renal interstitial fibrosis that is based on the recognition of all main tissue structures in the biopsy sample. This recognition step is useful in accurately estimating not only the interstitial fibrosis area, but also the relative health of the tissue based on the tissue structure appearances. The proposed quantification framework based on a comprehensive recognition of tissue structures is tested on a set of renal biopsy images with ground truth prepared in collaboration with an expert pathologist.

2. Proposed methodology

This section presents the proposed approach to the automatic quantification of interstitial fibrosis in renal biopsy images based on colour image segmentation and tissue structure identification. In this study, the biopsy samples collected were stained by the prevalently used Masson's trichrome (MT). Before the tissue structure segmentation step is carried out, a normalisation operation is applied on the image dataset to maximise the contrast between

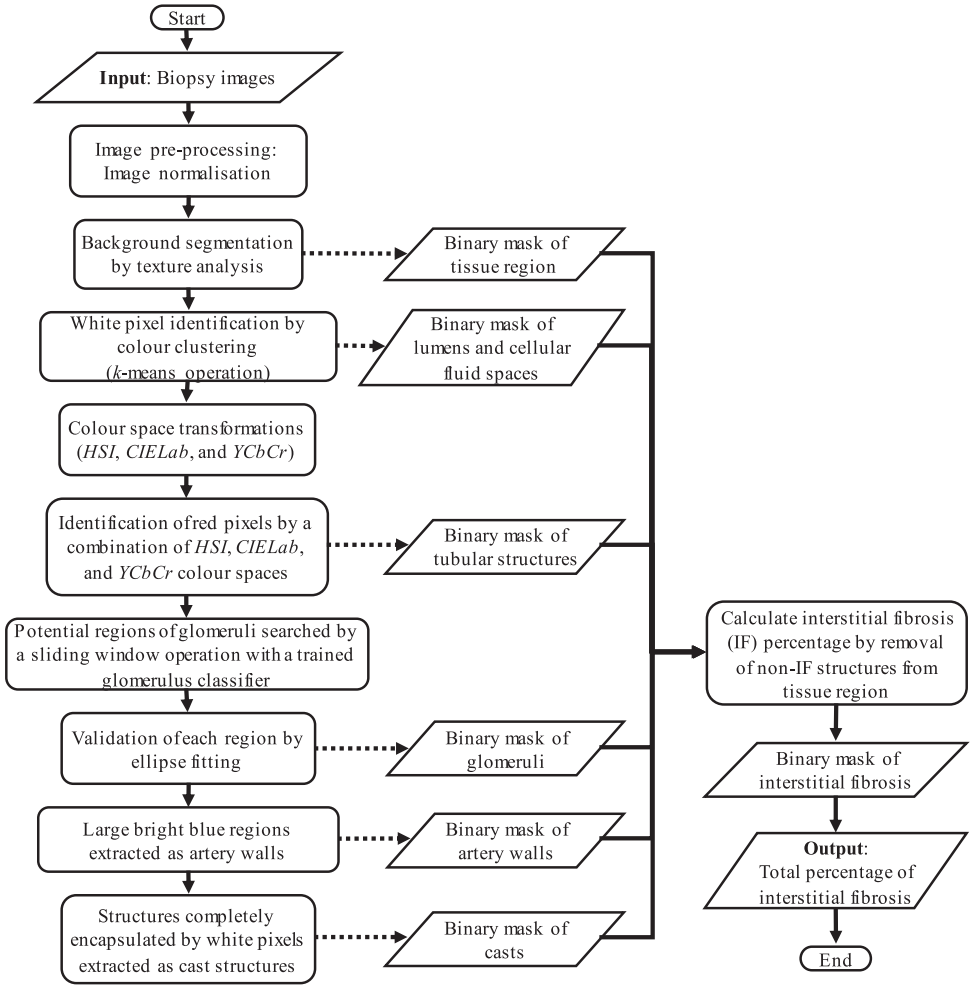


Fig. 1. Flow diagram of the proposed renal interstitial fibrosis quantification system, depicting the inputs and outputs of each stage.

the colours in the images. Subsequently, each of the kidney tissue structures are extracted by its colour and textural information obtained from the stained images, such as the tissue region of interest, the tubular structures, glomeruli, and artery walls. The flow diagram of the proposed quantification system is presented in Fig. 1.

2.1. Image normalisation

The captured images of biopsy samples often present non-uniformities due to variations in staining procedures, scanning equipment used, presence of folds in the samples, etc. These variations in the images will affect the accuracy of the algorithm in the detection of fibrotic area, while potentially introducing additional computational complexities in subsequent stages of processing. While it is nearly impossible account for all possible variations that can arise in a biopsy dataset, the contrast in each image was maximised to provide a basis for comparison between different images. To achieve this, the bottom and top 1% of all pixel values for each *RGB* histogram component of the image were coerced to the highest and lowest points of the new range. This step linearly expands the range of colour values present in the image to fill the maximum range of the colour histograms, thus facilitating the use of thresholds.

2.2. Background segmentation

To segment the *region of interest* (ROI) from the white background in the scanned image of the renal biopsy tissue, the conventional method is texture segmentation by edge detection. Compared to the smooth blank background region, the region of interest containing the tissue shows significantly greater information content, as shown in Fig. 2(a). Unfortunately, the uneven illumination in the imaging process often generates, to different degrees, a noisy speckled texture in the background region. To mitigate this problem, the colour edges instead of the intensity edges were detected based on colour invariance [32] in the image. In this case, the edges between two clearly separate colours are highlighted, regardless of the relative intensity difference. This method works well because the biopsy tissue region of interest has a clear contrast in colour with the white background. The colour invariant expressions are modelled under the assumption made by the Kubelka–Munk theory on the reflected light spectrum for coloured bodies [33]. The distinct states of different colours are discriminated using colour invariance by adopting a few assumptions on the imaging conditions. For inspecting biopsy images, the assumptions of matte, dull surfaces, planar objects, and an equal energy and uniform illumination was employed for the Kubelka–Munk model. These assumptions allow 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, as shown in Fig. 2(c).

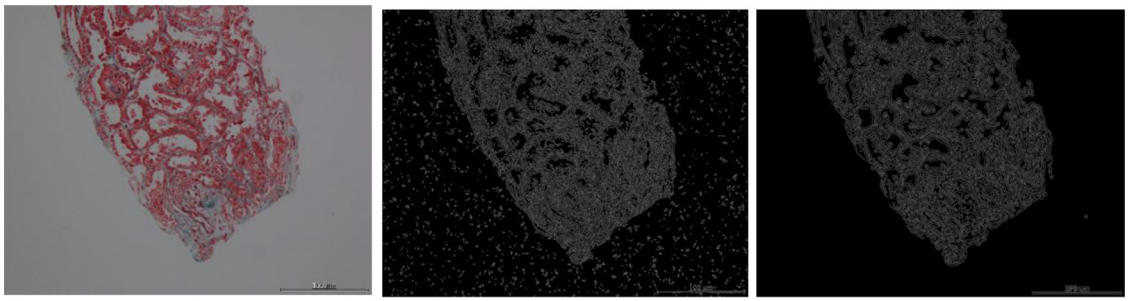


Fig. 2. The comparison of background segmentation methods (a) Original biopsy image sample (b) Edges in intensity of the image, in which noisy edges are detected (c) Edges in colour invariant image.

Table 1
List of commonly seen tissue structures and their methods for extraction, with example images in Fig. 3.

	Tissue structures	Extraction method
1.	Interstitial fibrosis	Deduction of biopsy area based on the exclusion of non-fibrosis structures (structures $3-8$).
2.	Inflammatory cells	Red structures smaller than 1000 pixels, distinguished from larger red tubular structures.
3.	Tubular structures	Large red structures derived from components from multiple colour spaces (*YCbCr, CIELab, HSI*).
4.	Lumens and cellular fluid spaces	White spaces obtained from *k*-means operation ($k=3$) with a conditional check for surrounding red structure (lumen spaces) and a size threshold of 1000 pixels (oedema fluid cavity).
5.	Healthy glomeruli	Biopsy regions displaying a glomerulus-like texture identified by a trained classifier, in addition to being elliptical in shape.
6.	Sclerosed glomeruli	Large elliptical blue structures based on the *Cb* component of the *YCbCr* colour space.
7.	Artery walls	Blue structures identified by the *Cb* component of the *YCbCr* colour space which are larger than 1000 pixels to accommodate for the variable appearance of arteries in biopsy samples.
8.	Casts	Structures enclosed by white spaces, typically on the interior of renal tubules.

2.3. Tissue structure extraction

To quantify the extent of interstitial fibrosis, the various types of structures present in the kidney sample must be identified. In examining Masson's trichrome (MT)-stained biopsy slides, the presence of only three main colours in the images that clearly defines the boundaries of different structures provides an advantage for segmentation by colour. Traditionally, the interstitial fibrosis percentage is calculated by counting and adding the number of pixels with the blue-green colour of the MT-stained fibrosis to be compared against the total tissue sample area. However, it should be noted that interstitial fibrosis is not the only possible structure that is stained blue-green. For example, protein casts that can be found in the renal tubules of diseased patients are also stained blue. Additionally, red inflammatory cells that are inherently found in the interstitial fibrosis should also be counted towards the final percentage interstitial fibrosis. Therefore, simple colour segmentation is insufficient to account for the various possible structures found in the renal biopsy sample, as proven in [34]. A knowledge-based colour segmentation approach is required to deduce a quantification result that more closely represent the actual extent of interstitial fibrosis.

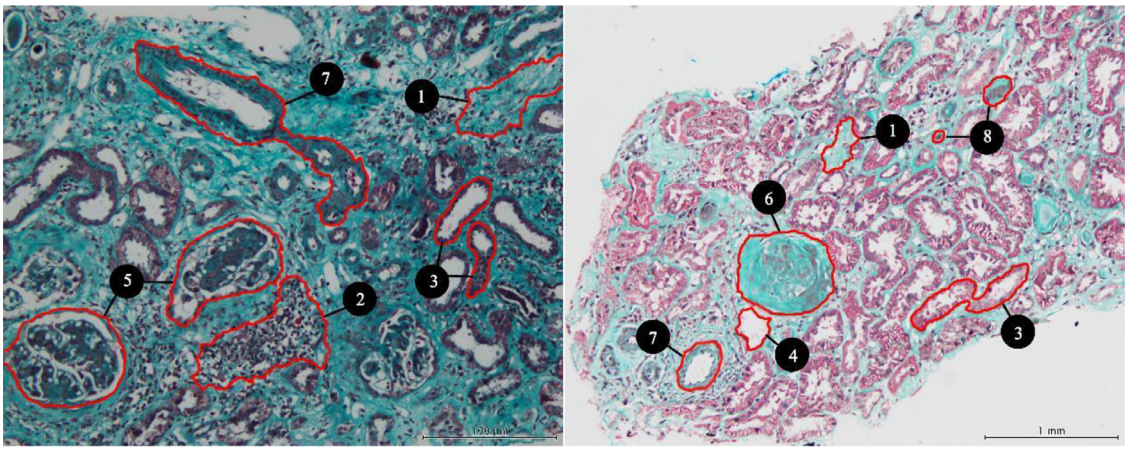
The various tissue structures that are to be extracted and removed from the segmented tissue region to obtain only the interstitial fibrosis structure are listed in Table 1, with the corresponding example images shown in Fig. 3. The methods employed for the identification of each structure are also detailed in the same table. It is worth noting that these are only the more commonly encountered structures in a renal biopsy and it is, by no means, an exhaustive list as there exists a wide variation in the human biology and pathology. It is exactly this reason that proves the superiority of a human evaluation, and the proposed system only attempts to provide an aide to the pathologists in their diagnosis by accounting for the statistically common structures in renal histopathology. A point to note is that the inflammatory cells (item 2 in Table 1) are only counted into the interstitial fibrosis area when they are lying on a background of interstitial fibrosis. The colour and textural operations involved in the segmentation of each of these structures are explained in the following sections.

2.3.1. Lumens and cellular fluid spaces

The lumens and cellular fluid spaces are the empty interior spaces of tubules and cells that are unstained by the Masson's trichrome during biopsy slide preparation. These white unstained regions were identified by employing a *k*-means clustering operation, with $k=3$ reflecting the three main colours (white, blue-green, and red) present in the image. White spaces in the tissue region can be classified into three common structures, namely the tubular lumen, cell vacuoles, and fluid cavity from oedema. Large white spaces obtained from the *k*-means image are identified as lumens when surrounded by red tubular walls, while the smaller ones are classified as cellular vacuoles in the interstitium to be considered towards the quantification of interstitial fibrosis. These small white spaces are selected from the white image by a size threshold of 1000 pixels at the objective magnification level of $10\times$ (about 130 square microns), which is the maximum size of a cell considered by pathologists. The less commonly encountered oedema fluid cavities are extracted as white spaces above the size of 1000 pixels, as these structures have virtually no distinguishing features other than its size compared to the cellular vacuoles.

2.3.2. Tubular structures

The next main structure to segment is the kidney tubular structures, which is stained red by the Masson's trichrome. Conventional colour detection in an image in the *RGB* colour space proves to be inherently flawed as a major weakness of the *RGB* colour model is its inability to separate the luminance and chrominance components, which makes it a poor choice for colour analysis and colour based recognition [35]. Therefore, it is proposed to use a combination of colour space components that correlate with the colour red to detect red pixels in the biopsy image. These colour space components include:



(a) (b)

Fig. 3. Examples of renal biopsy tissue with the commonly seen tissue structures labelled with corresponding numbers in Table 1.

- the *Cr* and *Cb* components of the *YCbCr* colour space,

$$C_b = 128 - \frac{37.945 \cdot R}{256} - \frac{74.494 \cdot G}{256} + \frac{112.439 \cdot B}{256} \tag{1}$$

$$C_r = 128 + \frac{112.439 \cdot R}{256} - \frac{94.154 \cdot G}{256} - \frac{18.285 \cdot B}{256} \tag{2}$$

- the *a* component of the *CIELab* colour space, and

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{3}$$

$$a^* = 500\left[\sqrt[3]{\frac{X}{X_o}} - \sqrt[3]{\frac{Y}{Y_o}}\right] \tag{4}$$

- the saturation component of the *HSI* colour space.

$$saturation = 1 - \frac{3}{R+G+B}\min(R, G, B) \tag{5}$$

The colour intensity for the red pixels in the biopsy image was derived from the cumulative effect of the colour space components previously listed, as given in (6)–(8). A detailed description of the derivation is provided in a previous research paper [34].

$$C_{b,invert} = 1 - C_b \tag{6}$$

$$C_{b,invert} = \begin{cases} 1 & for\ Cb_{invert} > Otsu\ threshold, \\ Cb_{invert} & otherwise \end{cases} \tag{7}$$

$$Red = C_r \times a^* \times saturation \times C_{b,invert} \tag{8}$$

The red pixels were then grouped into individual structures by a connected components operation. Before the non-tubular red structures (blood vessel walls and inflammatory cells) were differentiated from the tubular wall structures, a conditional check for the average size of the red connected components present in the image was carried out. An average size of red components in the image smaller than 400 pixels or about 50 square microns (threshold size determined experimentally based on 40 biopsy samples with annotated tissue structure regions) indicates the presence of a large number of inflammatory cells in the biopsy sample. In this case, the red pixels are determined by incorporating the colour invariance image to more accurately line out the colour edges in the dense inflammatory cell clusters. Otherwise, the red pixels are extracted simply by a dynamic Otsu threshold [36] of the red intensity image determined by (8). A connected components analysis was then applied on these red pixels to group them as individual structures before a size threshold of 1000 pixels was imposed on them. This size threshold corresponds to the minimum size of a red connected component to be considered a tubular wall structure. The connected components are morphologically filled to extract any interior white lumen spaces together with the tubular walls.

2.3.3. Glomeruli

The glomeruli are spherical tufts of capillaries in the kidney. 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. However, the presence of this blank space is dependent on the slide preparation process, just as the shape of the glomerulus. Therefore, the identification of glomeruli by the presence of Bowman's space is an unreliable method. A healthy glomerulus has a perforated appearance due to the dense clusters of cells in the glomerulus. This honeycomb-like appearance presents a distinct textural difference compared to its surrounding structures, and is therefore taken advantage of to identify potential glomerulus regions in the biopsy image.

The proposed steps towards the segmentation of glomerulus regions in the biopsy images are:

1. Train a classifier for detecting glomerulus regions with manually identified glomerulus images.
2. Perform a sliding window operation on the biopsy image to identify probable glomerulus areas. Specifically, the biopsy image is inspected by a window of 300×300 pixels that "slides" across the image with a 75% overlap to the previous window. Each window is tested for glomerulus by the classifier trained in Step 1.
3. Fit ellipses onto the identified probable areas to further refine the shape of the glomeruli, based on the assumption that glomeruli are approximately elliptical in shape. To account for the likely occurrence of close clusters of multiple glomeruli, multiple non-overlapping ellipses were allowed to be fitted onto each low intensity blob. The parameters for the ellipse fitting at the magnification level of $100\times$ are a minimum major axis length of 200 pixels and aspect ratio of 0.8, based on 20 manually annotated ground truth images.

The classifier employed to search out the potential glomerulus structures proposed in Step 1 works by examining the textural information in the local windowed region. The feature selected for training the classifier was the Local Binary Pattern (LBP) [37]. 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 matching the circular symmetry of glomerulus texture, and can thus reveal local micro-patterns such as edges, flat areas and spots in the texture image.

A comparison experiment for establishing the best classifier parameters to be included in the quantification system for the glomeruli detection algorithm is included in the Appendix. The parameters tested for the generation of LBP features were the rotational invariance and neighbourhood radius on the greyscale and red pixel intensity (calculated in (8)) images. The *Support Vector Machine* (SVM) classifier [38] with different kernel choices and the *Alternating Decision Trees* (ADTree) and its variant classifiers [39] were also compared. Based on the results, rotation invariant LBP features of the red intensity image and greyscale image with neighbourhood radii of 5 and 10 stands out as the most discriminatory feature set, while the SVM classifier with the kernel of the 4th order outperforms the other classifier variants. Finally, the confirmed glomeruli structures were then extracted and saved for subsequent processing.

2.3.4. Sclerosed glomeruli

As the kidney function deteriorates, it also causes scarring and hardening of the glomeruli structures in the kidney tissues. Once stained by Masson's trichrome, sclerosed glomeruli appear bright blue as they now consist of scarring tissues, similar to that of interstitial fibrosis but not included in its quantification. The extraction of this structure is carried out by identifying high intensity levels in the *Cb* component of the *YCbCr* colour space (Eq. (1)) with a size bigger than 1000 pixels in a roughly circular shape.

2.3.5. Artery walls

The renal artery is made up of smooth muscles which are stained red by the Masson's trichrome, and a hollow lumen at its centre. Due to its tubular nature, the shape it presents in a renal biopsy can vary between a circular and a long tubular structure. However, a distinguishing characteristic of artery walls is its surrounding uniformly-blue tissue structure. Therefore, a detection of high intensity level in the *Cb* component of the *YCbCr* colour space would highlight artery structures. A validation step of checking these regions against a size threshold of 1000 pixels would retain actual artery structures while eliminating noise and false targets.

2.3.6. Casts

In more severe cases of renal function deterioration, excess protein and blood are found in the patient's renal tubules. When these biopsy samples are taken and stained by Masson's trichrome, this pathological condition shows up in the biopsy sample in the form of urinary casts. Urinary casts can come in the form of protein casts and cellular casts like red blood cell casts. The casts are only found in the lumens of renal tubules. Since any urinary cast completely enclosed by tubular walls would already be previously extracted together with the renal tubules, this step would have to remove any leftover cast in renal tubules with fragmented walls. They can be extracted by simply removing structures surrounded by blank white space between the renal tubules and urinary casts.

2.4. Interstitial fibrosis quantification

Upon the exclusion of all the renal tissue structures that are not considered as interstitial fibrosis from the biopsy tissue sample, the final leftover region is the interstitial fibrotic region of interest to be quantified. The binary mask of the interstitial fibrosis area was obtained through the elimination of non-fibrosis structure regions from the region of interest mask segmented in Section 2.2. The amount of interstitial fibrosis in the biopsy sample is expressed as a percentage of the total tissue area segmented from the background, described by (**9**).

$$IF(\%) = \frac{ROI - nonIF}{ROI} \tag{9}$$

3. Experimental results

This section presents the experimental procedure in the testing of the proposed computerised renal interstitial fibrosis quantification system against a prepared dataset with the involvement of a group of skilled pathologists.

3.1. Experiments

Four experiments were carried out to gauge the feasibility of the proposed quantification system and the extent of variability in the visual quantification method. They are listed in Table 2.

3.2. Data acquisition

In this study, a total of five expert pathologists and 70 patients with kidney diseases from the Sultanah Aminah Hospital, Johor Bahru volunteered as test subjects. The skilled pathologists perform interstitial fibrosis quantification by visual evaluation on the kidney biopsy samples obtained from the patients. The equipment utilised in the acquisition of the biopsy image dataset and subsequent image processing results are as follow.

- Olympus BX53 system microscope to observe biopsy samples under an objective magnification of 10×. The magnification level was selected to emulate the pathologists' view in visual evaluation of the biopsy samples.
- Olympus DP21 camera for microscopes with a pixel resolution of 2448 × 1920 in JPEG format to digitise the biopsy samples.
- MATLAB software on an Intel® Core i5-3470 CPU with an 8GB RAM to perform computerised quantification on the digital biopsy images.

3.3. Image dataset

To evaluate the performance of the proposed interstitial fibrosis quantification system, 286 digital images of renal biopsy samples taken from the 70 patients involved were acquired. This biopsy sample set was selected to include the range diseased states in the kidney, from the healthy to acutely diseased, and are mostly transplant biopsy samples. The kidney tissues were stained by Masson's trichrome to facilitate the observation of fibrotic structures. These images were separated into several datasets to facilitate experimentation on the quantification of interstitial fibrosis in these biopsy images. The image datasets used for experiments are listed as follows.

I. 140 images of biopsy consisting of only tubular structures.
II. 146 images of biopsy with all possible tissue structures (renal tubules, glomeruli, arteries, etc.).
III. 40 manually segmented ground truth images highlighting the fibrotic areas were prepared and verified by a senior consultant pathologist with renal pathology subspecialty training. This dataset comprises of 20 images each from relatively healthy and diseased biopsy samples respectively.

3.4. Statistical analysis methods

Several performance indices were employed in this experiment to gauge the practicality of the computerised quantification system. These include the mean absolute error, accuracy, specificity

Table 2
The list of validation experiments conducted to test the performance of the computerised quantification system.

Experiment	Description	Image dataset	Aim
1	Computerised quantification of interstitial fibrosis by the identification of all tissue structures on biopsy images with manually annotated ground truth.	40 renal biopsy images (20 healthy, 20 diseased) each with ground truth of interstitial fibrosis regions manually segmented.	To obtain the quantitative accuracy of the computerised interstitial fibrosis quantification system.
2	Computerised quantification of interstitial fibrosis on a larger set of renal biopsy images.	A total of 286 renal biopsy images taken from 70 kidney patients.	To determine the consistency of the computerised interstitial fibrosis quantification system with the pathologists' quantification results.
3	Conventional histopathologic evaluation of interstitial fibrosis by five skilled pathologists on the same renal biopsy image sets used in Experiment 2.	A total of 286 renal biopsy images taken from 70 kidney patients.	To investigate the intra- and inter-observer variation in traditional quantification of interstitial fibrosis by visual evaluation.
4	Histopathologic evaluation of interstitial fibrosis aided by automated quantification results by five skilled pathologists on the renal biopsy image set containing all tissue structures.	A total of 146 renal biopsy images taken from 70 kidney patients.	To measure the effect of having the automated quantification results as a reference on the pathologists' visual quantification results.

and sensitivity. The *true positives* (*TP*) are pixels correctly classified as interstitial fibrosis, while *true negatives* (*TN*) are pixels correctly classified as non-interstitial fibrosis structures, and vice versa.

To gauge the amount of correlation between the different pathologists and the automatic quantification system, two main statistical methods were employed, namely the intraclass correlation coefficient [40] and kappa tests. The intraclass correlation coefficient (ICC) is used to measure the similarity between variables sharing the same metric and variance, while the kappa coefficient assesses the agreement between a fixed number of raters on categorical ratings with chance agreement considered. Two types of kappa coefficient are commonly used in machine learning, namely the Cohen's kappa [41] and Fleiss' kappa [42] for cases with only two observers and more than two observers respectively. For both the kappa coefficients, 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.

As a basis for subsequent discussion, the following categorical labels for the corresponding ranges of the kappa value introduced by Landis et al. [43] are used to convey the relative strength of agreement between the raters.

i. (<0.00) Poor agreement
ii. (0.00–0.20) Slight agreement
iii. (0.21–0.40) Fair agreement
iv. (0.41–0.60) Moderate agreement
v. (0.61–0.80) Substantial agreement
vi. (0.81–1.00) Almost Perfect agreement

3.5. Validation of computerised system performance against ground truth

The quantitative analysis on the performance of the proposed interstitial fibrosis quantification system based on ground truths is described in this section. Experiment 1 in Table 2 was carried out to provide a definitive measure of the accuracy of the proposed system. Therefore, the set of 40 biopsy images with ground truth annotated with the aid of a practicing pathologist was used, as was discussed in Section 3.3.

The binary mask for the pixels classified as interstitial fibrosis structure produced by the computerised quantification system for each image in the dataset was first obtained using the procedure described in Section 2.4. This mask was subsequently compared against the manually annotated binary mask for interstitial fibrosis from the corresponding ground truth image. A pixel-by-pixel evaluation between the two binary masks gives a resultant binary mask indicating the accuracy of the quantification system at the pixel level.

Table 3
Performance of computerised system in the detection of interstitial fibrosis against a 40-image ground truth dataset.

Performance index	Average value
Absolute error (percentage point)	9.1%
Accuracy	77.6%
Specificity	85.1%
Sensitivity	70.1%

Table 3 presents the average values for absolute error, accuracy, specificity, and sensitivity of the proposed computerised quantification system against the ground truth. An example of the resulting classification image of the computerised system and the ground truth image is shown in Fig. 4.

It can be shown from Table 3 that the mean absolute error is well below 10 percentage points, which corresponds to less than half of the interval of a Banff classification category [3]. This translates to a reasonably reliable quantification for the pathologists in the estimation of interstitial fibrosis extent in a biopsy image. An accuracy of close to 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 show in the biopsy. The lower sensitivity value is a necessary trade-off for a higher specificity. A much higher specificity of the system as compared to its sensitivity indicates its much higher capability to correctly identify non-interstitial fibrosis structures than interstitial fibrosis 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. Conversely, favouring a higher sensitivity value would lead to an increased false positive value in the system. This translates to an increase in falsely identified interstitial fibrosis regions, thereby discounting the usefulness of the quantification system as a medical aide.

3.6. Validation of computerised system performance against pathologists

This section provides a comparison between the performance of the proposed computerised quantification system and the conventional method of visual evaluation by skilled pathologists. Experiments 2 and 3 listed in Table 2 were carried out, where the

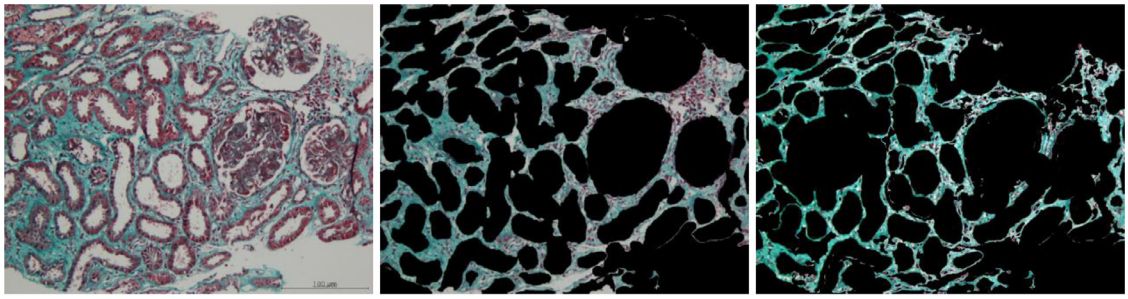


Fig. 4. (a) Sample renal biopsy (b) manually annotated ground truth of interstitial fibrosis structures (c) corresponding segmented interstitial fibrosis structures by the computerised quantification system.

Table 4
The inter-observer variability measures between different pathologists and between the automated quantification and pathologists' visual quantification in two datasets, with ICC = 1 and Fleiss' kappa = 1 in the case of perfect agreement.

Dataset	Pathologists without aide		Pathologists with aide		Computer system vs. Pathologists
	ICC	Fleiss' kappa	ICC	Fleiss' kappa	ICC
Tubular structures only (140 images)	0.6196	0.0349	–	–	0.7173
All structures included (146 images)	0.7082	0.1482	0.7460	0.2434	0.7034

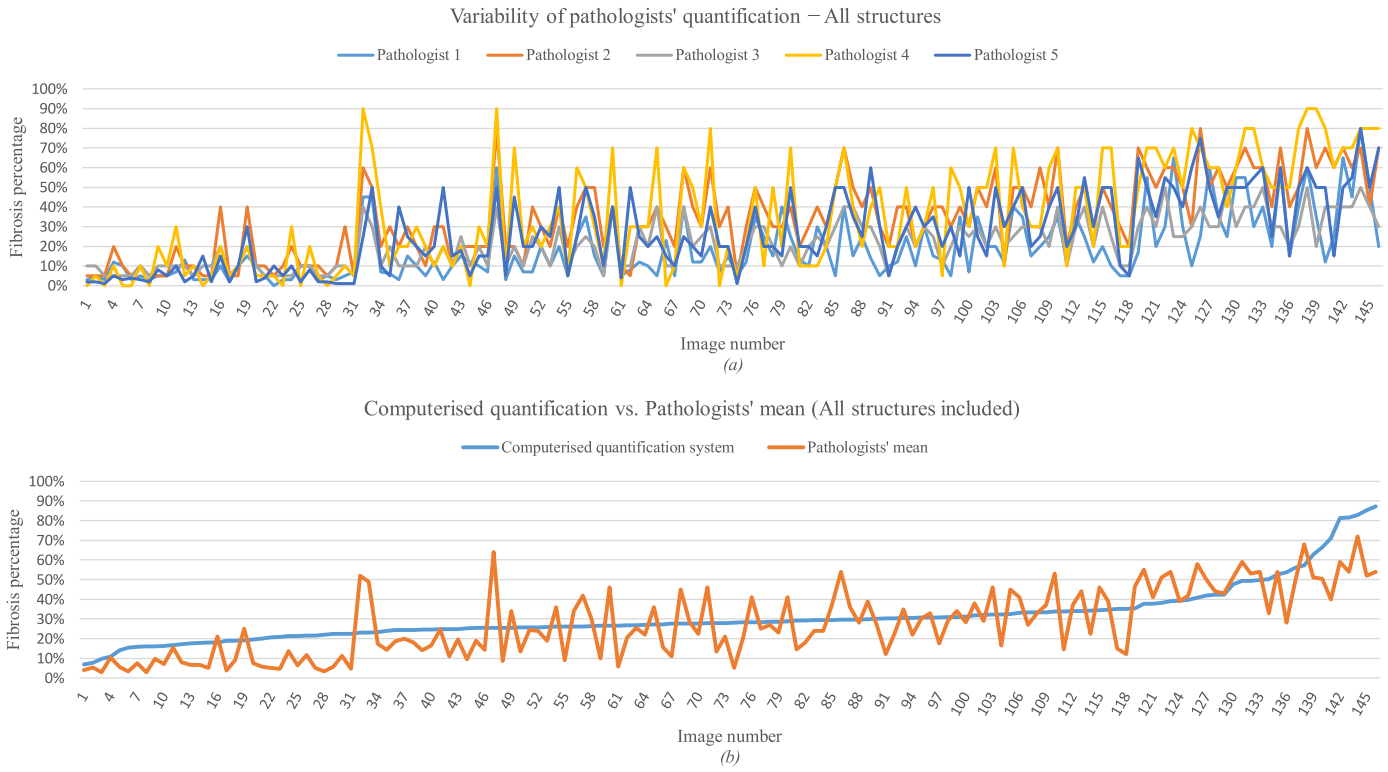


Fig. 5. The interstitial fibrosis quantification results on the biopsy dataset consisting of all structures demonstrating (a) the variability between the five pathologists, and (b) the comparison between the automated quantification and the average of two rounds of pathologists' scorings.

computer quantification system and the pathologists perform interstitial fibrosis quantification on the same image datasets. This method of comparison serves to compensate for the limitation presented by the relatively small size of the ground truth dataset, while offering a direct assessment of the correlation between automated and human quantification. However, the variability inherent in the conventional quantification method by visual evaluation of pathologists is a well-explored topic in the literature [4,5]. Therefore, the exact amount of pathologist variability is investigated in this section in order to gauge the validity of the agreement between the automated quantification and the average pathologist quantification. The pathologists' intra-observer variability on dataset (i) in Section 3.3 has been investigated in an earlier research [44], with the Cohen's kappa values ranging between 0.23 to 0.31 indicating a fair agreement in the pathologists' individual quantifications.

The unpredictability in the quantification result by visual evaluation is not only limited to that within the individual, but also to that between different pathologists as well, if not more prominent. The variability between the five independent pathologists was assessed by the intra-class correlation coefficient and the Fleiss kappa test. For the Fleiss kappa test, the pathologists' quan-

tification percentage results were categorised to the nearest 10%. The same 286-image dataset was used in this investigation.

Table 4 presents the results for the inter-observer variability tests. It is observed that the Fleiss' kappa values for both datasets between the different pathologists without the computerised aide are drastically low, corresponding to only a very slight agreement based on the categorisation by Landis et al [43]. This suggests that while the pathologists may agree in different relative scales of quantification as shown by the intraclass correlation coefficient, their actual scores for the same images vary significantly enough to not produce a quantification score that conveys the same diagnosis to other pathologists.

Fig. 5(a) demonstrates the correlation between the scorings of five independent pathologists on dataset (ii) in Section 3.3. The most apparent observation from the graph is that the pathologists have a consensus on the relative extent of fibrosis progression, as indicated by the matching peaks and valleys of the scores by different pathologists. In other words, the pathologists may agree on the relative severity of fibrosis present in the biopsy images, despite each having a different perception on the extremes. For example, pathologist 1 is more conservative in scoring compared to pathologist 4, as shown by the lower variation in pathologist 1′s scores.

Another observation from the comparison between the quantification for the two different datasets is the greater agreement between the pathologists in the estimation of the interstitial fibrosis amount in biopsy images with all types of tissue structures present. This is proven by the slightly higher intraclass correlation coefficient of 0.71 for the dataset with all structures as opposed to 0.62 for the tubules-only dataset, as shown in Table 4. One possible rationale behind this observation is the clearer distinction between healthy and diseased renal biopsy samples in the dataset with all structures present. In other words, most of these images show either very small or large amount of interstitial fibrosis in the kidney samples, as can be seen from the pathologist quantification results in Fig. 5(a). The human observer produces better estimations of extreme cases than the more ambiguous average cases.

3.7. Reliability of computerised quantification as a second opinion

One of the primary purposes of the proposed computerised quantification system is to provide a reasonable basis for discussion between pathologists on the diagnosis of renal interstitial fibrosis 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 practicing pathologists. A previous work by Farris et al. [45] has demonstrated the variability among pathologists from different centres and has suggested the use of an automated assessment as a tool for standardisation. Experiment 4 (Table 2) in this study furthers their work by examining the feasibility of having an automated quantification system as a second opinion in clinical diagnosis.

Based on the value of intraclass correlation coefficients for the comparison between the computer system and pathologists in Table 4, a fairly high amount of consistency was observed for both the datasets. It is also proven in Fig. 5(b) that the automated quantification result mimics the trend of the pathologists' visual quantification result almost perfectly, despite the presence of a small linear shift. The degree of quantification variation between different pathologists is on a comparable level to that between the computer and pathologists, as demonstrated by the under 0.1 difference in ICC values between the two types of correlations investigated (0.62 vs. 0.72 and 0.71 vs. 0.70 for the two datasets respectively). In fact, a higher computer-pathologist ICC value in the case of the tubules-only dataset suggests that the automated system has a higher agreement with the average pathologist than the pathologists do among themselves.

The impact of the automated quantification as an aide to the pathologists' visual quantification was investigated in Experiment 4 and measured by the ICC and Fleiss' kappa test values recorded in Table 4. 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 0.71 to 0.75) and the Fleiss' kappa (increase from 0.15 to 0.24) tests carried out. The degree of increased correlation is also reflected in the rise of one agreement category of the Fleiss' kappa from a slight to fair agreement.

In order to provide a direct assessment of the reliability of the proposed system in the diagnostic process of chronic kidney diseases, the quantification results of interstitial fibrosis were classified according to the Banff criteria [3]. The three grades of interstitial fibrosis as per the Banff classification are:

I. mild (interstitial fibrosis $< 25\%$ of cortical area),
II. moderate (interstitial fibrosis 26–50% of cortical area), and
III. severe (interstitial fibrosis $> 50\%$ of cortical area).

The intraclass correlation and Cohen's kappa statistics were calculated to compute the diagnosis consistency between the computer system and the average of the five pathologists. The intraclass correlation between the five pathologists' diagnoses was also computed, with the results compiled in Table 5.

Referring to Table 5, the computer-pathologists ICC values are consistently higher than the inter-pathologist ICC for both datasets. This result supports the previous evidence that the automated quantification system can produce a diagnosis which is in accordance with the average expert pathologist. Furthermore, the degree of correlation exhibited has shown to be consistently higher than the agreement between different pathologists. On the other hand, the Cohen's kappa values have also confirmed a fair agreement between the computer and pathologists' quantification with values of 0.25 and 0.41 respectively. In fact, the higher correlation for the image dataset with all possible tissue structures demonstrates the robustness of the computer system to account for the different types of structures in a biopsy image towards interstitial fibrosis quantification.

The effect of having a computer aide in the visual quantification by the pathologists classified according to the Banff criteria was also evaluated by the ICC and Fleiss' kappa values given in Table 5. The results have again demonstrated a positive effect of the aide on the correlation between the pathologists, with a rise in both the ICC (from 0.58 to 0.65) and Fleiss' kappa values (from 0.31 to 0.43) when contrasted against the pathologists' correlation without the aide. Furthermore, the Fleiss' kappa value has exhibited a rise in agreement from a fair to moderate agreement category. This implies the feasibility of the automated quantification system performing as an independent observer as proficient as the average pathologist.

4. Discussion

The main contributions of this paper are listed as follows.

- The proposal of an interstitial fibrosis quantification system that uses a combination of colour spaces in the segmentation of all kidney tissue structures in the biopsy image. Taking advantage of the colour cues presented by stained biopsy samples, the proposed method explores the use of multiple colour spaces to expand the range of information available for tissue structure extraction. All the kidney tissue structures in the biopsy image were deduced by a set of defined knowledge-based rules. Differing from other fibrosis segmentation methods, the proposed

Table 5
The variability measures between the automated quantification and pathologists' visual quantification and between different pathologists in terms of the Banff categories, with ICC = 1 and Cohen's kappa = 1 in the case of perfect agreement.

Dataset	Computer system vs. Pathologists		Pathologists without aide		Pathologists with aide	
	ICC	Cohen's kappa	ICC	Fleiss' kappa	ICC	Fleiss' kappa
Tubular structures only (140 images)	0.6212	0.2526	0.4452	0.1532	–	–
All structures included (146 images)	0.5825	0.4088	0.5787	0.3098	0.6472	0.4345

method does not only identify the pixel colour indicative of the interstitial fibrosis structure as highlighted by the chemical stain [16,22,23,25,46], but also the rest of the structures found in the biopsy tissue. This is to accommodate for the inclusion of the inflammatory cells in the final quantification value of interstitial fibrosis, following the assumption made by pathologists in their visual evaluation of the biopsy images.

- The validation of a glomerulus structure detector based on the image textural feature of Local Binary Patterns was presented. The glomerulus detector functions by the extraction of LBP features in a sliding window across the image to search for potential glomerulus regions through a trained LBP feature classifier. A total of 11 variants of the LBP features and five variants each for the SVM and LADTree classifiers were tested in a comparative experiment to determine the best performing glomerulus classifier. The classifier used was the SVM classifier employing rotation invariant LBP features of the red intensity image and greyscale image with neighbourhood radii of 5 and 10, attaining a 94% accuracy on the training image set.
- A ground truth image dataset denoting interstitial fibrosis regions was carefully prepared and annotated with the help of experts for this study which has allowed a more accurate measure of the performance of the proposed quantification system. The reason behind the lack of an established ground truth image set in the literature on interstitial fibrosis quantification is 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 interstitial fibrosis regions 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.

The feasibility of the proposed interstitial fibrosis quantification system has been evaluated by the series of experiments involving experienced pathologists in this study. The purpose of these experiments is twofold: they evaluate the performance of the computerised quantification system against a smaller set of images with annotated interstitial fibrosis regions and measure the degree of correlation between the computerised quantification and the pathologists' visual evaluation result. Additionally, the correlation and variability between the pathologists were also investigated to provide a basis of comparison. Based on the results obtained, it has been proven that pathologists' variability is a real issue in the diagnosis of renal disease as the same pathologist's judgment may change as much as 12% on average in a repeated quantification. This corroborates with results from related studies done on human observer variability [4–6]. However, all pathologists are still in general agreement on the relative severity of interstitial fibrosis extent in the biopsy samples as shown by the ICC values. Therefore, the computerised quantification tool attempts to provide a reference scale that can be examined objectively by different users, not only in the practising medical area, but the biomedical research field in general. Considering the experimental results of the correlation between the automated and pathologists' quantification, the computerised estimation of the interstitial fibrosis amount is as reliable as the average independent human observer. An added advantage over the human observer is the complete reproducibility in the quantification by the computer system.

In consideration of the wide variation seen in the human biopsy specimens, trade-offs in the decision-making process are unavoidable. This is especially evident in the identification process for the red pixels and its connected components. For example, the conditional check on the average size of red structures implemented in Section 2.3.2 is necessary to identify the more severely diseased biopsy specimens through the inference on the density of inflammatory cells present. This allows a modified method for red structures detection which minimises the false segmentation of inflammatory cells as large tubular structures to be executed. A minor sacrifice was made on the accuracy of interstitial fibrosis segmentation (about 3%) when compared against the ground truth dataset during experimentation. However, the implementation of this step has significantly raised the correlation between the computerised quantification and the average pathologist's quantification from 0.4847 to 0.7034 (as recorded in Table 4) for the 146-image dataset containing all possible tissue structures.

Nevertheless, several limitations in the system exist and remain to be developed upon. For instance, the relatively small size of the dataset used in this study is a limiting factor to the inclusiveness and robustness to the wide variation possible in the human pathology. However, the development of a graphical user interface tool for the pathologists to make amendments and provide feedback to the system will allow improvements to be made in future versions of the system. This also follows for the expansion of the feature set of the glomerulus cases, in which unidentified glomerulus instances can be added into the database by pathologists. Other issues include the use of images of sections of biopsies as opposed to whole slide images, and the magnification-specific parameters used in this proposed system. The processing of the whole area of a biopsy can be carried out block-by-block with cumulative quantification values with virtually identical results, while parameter values can be inter- or extrapolated to suit the magnification level of the image used. Biopsy specimens containing the occasional severe artefact are manually removed from the dataset, as they are the result of human error which would compromise the accuracy of quantification by visual evaluation.

5. Conclusion

This paper has presented a computerised tool for the automatic quantification of renal interstitial fibrosis in digital biopsy images. The proposed approach is based on the identification and segmentation of all commonly seen tissue structures in renal biopsies in order to achieve a more accurate calculation of the amount of interstitial fibrosis structures in the biopsies. Upon evaluation against the ground truth dataset, the system achieved an average error of within nine percentage points and an average pixel accuracy of 78%. Experiments testing the effectiveness of the system as a di-

agnostic aide to the pathologists have revealed a rise in correlation between different pathologists by as much as one category of agreement in the kappa test, proving the capability of the system as a reference standard in interstitial fibrosis quantification. In a clinical setting, the automated quantification of interstitial fibrosis will help to reduce the inter-observer variability between practising pathologists. Thus, this contributes to a more objective estimation of the amount of fibrosis in renal biopsy tissues.

Acknowledgements

The authors would like to thank the pathologists from Sultanah Aminah Hospital, Johor Bahru for their participation in the visual evaluation experiments. This research was supported by the Malaysian Ministry of Higher Education Fundamental Research Grant Scheme FRGS/2/2014/TK03/MUSM/02/1.

Appendix

The different approaches to recognising glomerulus structures in biopsy images are contrasted in this section in order to select the best algorithm for glomerulus detection and subsequent segmentation. A set of 528 individual glomerulus images and 514 images of non-glomerulus region were compiled from the biopsy images collected from the 70 patients to make up the glomerulus classifier training image set.

The 10-fold cross validation technique was used to assess the performance of the various glomerulus structure classifiers trained with the different parameters involving the Local Binary Patterns (LBP) feature set calculation. Based on the results shown in Table A.1, rotation invariant LBP features of the red intensity image and greyscale image with neighbourhood radii of 5 and 10 stands out as the most discriminatory feature set, while the SVM classifier with the kernel of the 4th order outperforms the other classifier variants. This outcome proved the effectiveness of the LBP features in capturing the textural information unique to the glomerulus structure, as initially suspected, due to the circular manner by which the features are calculated. This method fits with the elliptical shape of the glomeruli, thus allowing a more effective mapping of the distribution of intensity in the structure. The best performing classifier was integrated in the quantification system algorithm for glomerulus segmentation.

Glossary of terms

- Colour spaces

  The description of the range of colours that can be reproduced in a digital image, described by a mathematical relationship that correlates particular colours to physical ink colours or illuminated displays on digital display devices. Example colour spaces are *RGB, YCbCr, CIEXYZ, CIELab*, etc.

- Colour histogram

  The colours in a digital image are represented by real integers. The histogram representing the distribution of colours in an image by counting the number of occurrences of each number (colour). In other words, it shows the number of pixels corresponding to each colour in a digital image.

- Connected components operation

  An algorithm for detecting connected groups of pixels in an image, usually a binary image. A pixel is considered to be connected to another pixel of similar value if it is adjacent to the pixel.

- Dynamic Otsu threshold

  A dynamically determined threshold for a greyscale image to reduce it to a binary image. This method selects the threshold by maximising the variance between the "high-valued" and "low-valued" classes.

- *k*-means clustering

  An iterative clustering method which groups a set of data points into a predetermined *k* number of separate clusters. The assignment of each data point into a cluster is based on the maximisation of the inter-cluster distance.

- Luminance and chrominance

  The luminance is the brightness component while the chrominance is the colour component of an image.

Table A.1
Comparative classification results for the different classifiers and the Local Binary Pattern features used as training data to determine the best performing glomeruli detection classifier.

Classifier	*Features*										
	Red intensity					Greyscale					Red intensity + Greyscale
	No RI	With RI	With RI, radius 5	With RI, radius 10	With RI, radius 5 and 10	No RI	With RI	With RI, radius 5	With RI, radius 10	With RI, radius 5 and 10	With RI, radius 5 and 10 for red AND greyscale
	SVM kernels										
Linear	69.6%	70.1%	73.1%	74.7%	81.2%	68.6%	68.7%	70.8%	55.9%	79.9%	86.8%
Gaussian	70.9%	71.1%	75.7%	76.9%	83.8%	76.5%	70.9%	72.7%	57.8%	81.8%	89.2%
Polynomial order 2	71.8%	71.3%	80.4%	80.1%	85.9%	80.3%	73.8%	73.9%	61.4%	86.0%	91.9%
Polynomial order 3	74.0%	73.2%	83.4%	84.4%	87.2%	84.6%	77.1%	76.4%	68.8%	88.8%	93.1%
Polynomial order 4	77.7%	74.2%	84.1%	86.4%	87.4%	85.6%	78.9%	77.4%	72.0%	89.1%	***94.0%***
	ADTree										
Univariate	68.4%	67.5%	75.8%	82.3%	82.2%	66.6%	68.9%	70.8%	54.6%	70.9%	82.2%
Multivariate	80.1%	75.5%	83.5%	77.0%	77.7%	73.9%	80.3%	79.2%	62.3%	89.5%	92.5%
Sparse	78.8%	75.1%	83.4%	86.2%	86.4%	78.8%	79.9%	78.8%	71.1%	88.7%	92.8%
Lasso	78.8%	75.6%	83.0%	85.2%	86.1%	80.0%	80.1%	78.5%	71.8%	89.3%	92.6%
Elastic Net	80.1%	76.1%	83.2%	84.4%	85.2%	79.9%	80.0%	77.6%	71.4%	89.0%	93.2%

- Sliding window operation

  An operation in which the image is inspected one small region at a time. This region is typically set up as a rectangular window of fixed size and is systematically moved over the entire image.

- Statistical classifier

  An algorithm that predicts the category in which a new observation belongs to, based on a training set of observations with their corresponding labelled categories.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2017.12.004.

References

[1] V. Jha, G. Garcia-Garcia, K. Iseki, Z. Li, S. Naicker, B. Plattner, et al., Chronic kidney disease: global dimension and perspectives, Lancet 382 (2013) 260–272.
[2] B. Afzali, S. Jayawardene, D. Goldsmith, Diagnostic tests in chronic kidney disease, Kidney Dis. (2007) 1.
[3] K. Solez, R. Colvin, L. Racusen, M. Haas, B. Sis, M. Mengel, et al., Banff 07 classification of renal allograft pathology: updates and future directions, Am. J. Transplantation 8 (2008) 753–760.
[4] N. Marcussen, T. Olsen, H. Benediktsson, L. Racusen, K. Solez, Reproducibility of the Banff classification of renal allograft pathology: Inter-and intraobserver variation, Transplantation 60 (1995) 1083–1089.
[5] P.N. Furness, N. Taub, C. o. E.R.T.P.A.P. Project, International variation in the interpretation of renal transplant biopsies: report of the CERTPAP Project, Kidney Int. 60 (2001) 1998–2012.
[6] D. Serón, F. Moreso, X. Fulladosa, M. Hueso, M. Carrera, J.M. Grinyó, Reliability of chronic allograft nephropathy diagnosis in sequential protocol biopsies, Kidney Int. 61 (2002) 727–733.
[7] F. Bolger, G. Wright, Assessing the quality of expert judgment: issues and analysis, Decis. Support Syst. 11 (1994) 1–24 1.
[8] J.M. Gauch, C.W. Hsia, Comparison of three-color image segmentation algorithms in four color spaces, Appl. Opt. Sci. Eng. (1992) 1168–1181.
[9] A.P. Dhawan, A. Sim, Segmentation of images of skin lesions using color and texture information of surface pigmentation, Comput. Med. Imaging Graph. 16 (1992) 163–177.
[10] I. Giotis, N. Molders, S. Land, M. Biehl, M.F. Jonkman, N. Petkov, MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images, Expert Syst. Appl. 42 (2015) 6578–6585.
[11] D. Jiwen, L. Jing, F. Aifang, L. Huiming, Automatic segmentation for ovarian cancer immunohistochemical image based on YUV color space, in: Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on, 2010, pp. 1–4.
[12] K.Z. Mao, P. Zhao, P.-H. Tan, Supervised learning-based cell image segmentation for p53 immunohistochemistry, IEEE Trans. Biomed. Eng. 53 (2006) 1153–1163.
[13] S.H. Rezatofighi, H. Soltanian-Zadeh, Automatic recognition of five types of white blood cells in peripheral blood, Comput. Med. Imaging Graph. 35 (2011) 333–343.
[14] A. Sadr, M. Jahed, P. Salehian, A. Eslami, Leukocyte's nucleus segmentation using active contour in YCbCr colour space, in: Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on, 2010, pp. 257–260.
[15] X. Wang, B. Zhang, Z. Yang, H. Wang, D. Zhang, Statistical analysis of tongue images for feature extraction and diagnostics, IEEE Trans. Image Process. 22 (2013) 5336–5347.
[16] M. Masseroli, T. Caballero, F. O'Valle, R.M.D. Moral, A. Pérez-Milena, R.G.D. Moral, Automatic quantification of liver fibrosis: design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis, J. Hepatol. 32 (2000) 453–464.
[17] K. Meejaroen, C. Chaweechan, W. Khodsiri, V. Khu-smith, U. Watchareeruetai, P. Sornmagura, et al., Detection of fibrosis in liver biopsy images by using Bayesian classifier, in: Knowledge and Smart Technology (KST), 2015 7th International Conference on, 2015, pp. 184–189.
[18] L.B. Sant'Anna, N. Sant'Anna, O. Parolini, Application of computer-assisted image analysis for identifying and quantifying liver fibrosis in an experimental model, J. Comput. Interdiscip. Sci. 2 (2011) 139–148.
[19] K. Marten, V. Dicken, C. Kneitz, M. Hoehmann, W. Kenn, D. Hahn, et al., Computer-assisted quantification of interstitial lung disease associated with rheumatoid arthritis: preliminary technical validation, Eur. J. Radiol. 72 (2009) 278–283.
[20] M.D. Taylor, J.R. Roberts, A.F. Hubbs, M.J. Reasor, J.M. Antonini, Quantitative image analysis of drug-induced lung fibrosis using laser scanning confocal microscopy, Toxicol. Sci. 67 (2002) 295–302.
[21] D. Daunoravicius, J. Besusparis, E. Zurauskas, A. Laurinaviciene, D. Bironaite, S. Pankuweit, et al., Quantification of myocardial fibrosis by digital image analysis and interactive stereology, Diagn. Pathol. 9 (2014) 114.
[22] A.M. Hadi, K.T. Mouchaers, I. Schalij, K. Grunberg, G.A. Meijer, A. Vonk-Noordegraaf, et al., Rapid quantification of myocardial fibrosis: a new macro-based automated analysis, Cell. Oncol. 34 (2011) 343–354.
[23] M. Masseroli, F. O'Valle, M. Andujar, C. Ramirez, M. Gomez-Morales, J. de Dios Luna, et al., Design and validation of a new image analysis method for automatic quantification of interstitial fibrosis and glomerular morphometry, Lab. Invest. 78 (1998) 511–522.
[24] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population mixture models, Pattern Recognit. 25 (1992) 1231–1240.
[25] A. Servais, V. Meas-Yedid, M. Buchler, E. Morelon, J.-C. Olivo-Marin, Y. Lebranchu, et al., Quantification of interstitial fibrosis by image analysis on routine renal biopsy in patients receiving cyclosporine, Transplantation 84 (2007) 1595–1601.
[26] F. Moreso, D. Seron, J. Vitria, J.M. Grinyo, F.M. Colome-Serra, N. Pares, et al., Quantification of interstitial chronic renal damage by means of texture analysis, Kidney Int. 46 (1994) 1721–1727.
[27] M. Klapczynski, G.D. Gagne, S.J. Morgan, K.J. Larson, B.E. LeRoy, E.A. Blomme, et al., Computer-assisted imaging algorithms facilitate histomorphometric quantification of kidney damage in rodent renal failure models, J. Pathol. Inf. 3 (2012) 20.
[28] V. Meas-Yedid, A. Servais, L.-H. Noël, C. Panterne, P. Landais, N. Hervé, et al., New computerized color image analysis for the quantification of interstitial fibrosis in renal transplantation, Transplantation 92 (2011) 890–899.
[29] A. Bohle, S. Mackensen-Haen, H. von Gise, Significance of tubulointerstitial changes in the renal cortex for the excretory function and concentration ability of the kidney: a morphometric contribution, Am. J. Nephrol. 7 (1987) 421–433.
[30] F. Strutz, G. Müller, On the progression of chronic renal disease, Nephron 69 (1995) 371–379.
[31] A. Bohle, The pathogenesis of chronic renal failure in the primary glomerulopathies, renal vasculopathies, and chronic interstitial nephritides, Kidney Int. (1996) S2–S9.
[32] J.-M. Geusebroek, R. Van den Boomgaard, A.W.M. Smeulders, H. Geerts, Color invariance, Pattern Anal. Mach. Intell. IEEE Trans. 23 (2001) 1338–1350.
[33] P. Kubelka, New contributions to the optics of intensely light-scattering materials. Part I, JOSA 38 (1948) 448.
[34] W.K. Tey, Y.C. Kuang, J.J. Khoo, M.P.-L. Ooi, 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.
[35] A. Cheddad, J. Condell, K. Curran, P. McKevitt, A new colour space for skin tone detection, in: Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp. 497–500.
[36] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (1975) 23–27.
[37] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.
[38] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[39] H.K. Sok, M.P.-L. Ooi, Y.C. Kuang, S. Demidenko, Multivariate alternating decision trees, Pattern Recognit. 50 (2016) 195–209.
[40] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol. Bull. 86 (1979) 420.
[41] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.
[42] J.L. Fleiss, Measuring nominal scale agreement among many raters, Psychol. Bull. 76 (1971) 378.
[43] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.
[44] J.J. Khoo, W.K. Tey, V. Tan, S.W. Peter, Y.C. Kuang, M.P.L. Ooi, Visual quantification of renal interstitial fibrosis: inter and intra-observer variations, Pathology 49 (2017) S130.
[45] A. Farris, S. Chan, J. Climenhaga, B. Adam, C. Bellamy, D. Seron, et al., Banff fibrosis study: multicenter visual assessment and computerized analysis of interstitial fibrosis in kidney biopsies, Am. J. Transplantation 14 (2014) 897–907.
[46] A.L. Lazzarini, R.A. Levine, R.J. Ploutz-Snyder, S.O. Sanderson, Advances in digital quantification technique enhance discrimination between mild and advanced liver fibrosis in chronic hepatitis C, Liver Int. 25 (2005) 1142–1149.
[47] M.G. Tsipouras, N. Giannakeas, A.T. Tzallas, Z.E. Tsianou, P. Manousou, A. Hall, et al., A methodology for automated CPA extraction using liver biopsy image analysis and machine learning techniques, Comput. Methods Programs Biomed. 140 (2017) 61–68.