Imperial College London

Department of Bioengineering

Methods for Analysing Endothelial Cell Shape and Behaviour in Relation to the Focal Nature of Atherosclerosis

Saadia Iftikhar

May 2011

Supervised by: Anil A. Bharath

Peter Cashman

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Bioengineering of Imperial College London
and the Diploma of Imperial College London

Abstract

The aim of this thesis is to develop automated methods for the analysis of the spatial patterns, and the functional behaviour of endothelial cells, viewed under microscopy, with applications to the understanding of atherosclerosis.

Initially, a radial search approach to segmentation was attempted in order to trace the cell and nuclei boundaries using a maximum likelihood algorithm; it was found inadequate to detect the weak cell boundaries present in the available data. A parametric cell shape model was then introduced to fit an equivalent ellipse to the cell boundary by matching phase-invariant orientation fields of the image and a candidate cell shape. This approach succeeded on good quality images, but failed on images with weak cell boundaries. Finally, a support vector machines based method, relying on a rich set of visual features, and a small but high quality training dataset, was found to work well on large numbers of cells even in the presence of strong intensity variations and imaging noise.

Using the segmentation results, several standard shear-stress dependent parameters of cell morphology were studied, and evidence for similar behaviour in some cell shape parameters was obtained in *in-vivo* cells and their nuclei. Nuclear and cell orientations around immature and mature aortas were broadly similar, suggesting that the pattern of flow direction near the wall stayed approximately constant with age. The relation was less strong for the cell and nuclear length-to-width ratios.

Two novel shape analysis approaches were attempted to find other properties of cell shape which could be used to annotate or characterise patterns, since a wide variability in cell and nuclear shapes was observed which did not appear to fit the standard parameterisations. Although no firm conclusions can yet be drawn, the work lays the foundation for future studies of cell morphology.

To draw inferences about patterns in the functional response of cells to flow,

which may play a role in the progression of disease, single-cell analysis was performed using calcium sensitive florescence probes. Calcium transient rates were found to change with flow, but more importantly, local patterns of synchronisation in multi-cellular groups were discernable and appear to change with flow. The patterns suggest a new functional mechanism in flow-mediation of cell-cell calcium signalling.

Declaration of Originality

The work presented in this thesis was carried out between April 2007 and March 2010 in the Department of Bioengineering at Imperial College London under the supervision of Dr Anil A. Bharath and Dr Peter Cashman. The material herein is original and has not been submitted for a degree or diploma at any other university.

Image data were provided by the Physiological Flow research group, Department of Bioengineering, Imperial College London. More specifically, the rabbit data, described in Section 1.7.1, is provided by Dr Andrew R. Bond and the *in-vitro* temporal data, described in Section 1.7.2, is provided by Dr Nicolas Foin. The author has developed different algorithms for automated analysis based on code developed by herself and standard and modified libraries of *Matlab*, unless otherwise stated.

Saadia Iftikhar

Publications

Publications that have been produced based on the material in this thesis are as follows:

- Andrew R. Bond, *et al.*, "Morphological evidence for a change in the pattern of aortic wall shear stress with age" Arteriosclerosis, Thrombosis, and Vascular Biology, 2010. Some results of Chapter 4 are presented in this paper [25].

- S. Iftikhar, *et al.*, "Segmentation of Endothelial Cell Boundaries of Rabbit Aortic Images using a Machine Learning Approach", accepted for International Journal of Biomedical Imaging, 2011. Some results of Section 3.7 are presented in this paper.

- S. Iftikhar, *et al.*, "Calcium spikes form intricate inter-cellular patterns that are affected by flow", in writing.

- S. Iftikhar, *et al.*, "Analysis of Aortic Endothelial Cells and their Nuclei to understand Flow and Disease Patterns", gave a talk in bioengineering 2010, conference.

- S. Iftikhar, *et al.*, "Cellular Ca^{2+} oscillations are synchronized and react to flow, in Single Cell Analysis Summit, San Diego, USA, 2010.

- S. Iftikhar, and A. A. Bharath, "Cell boundary analysis using radial search for dual staining techniques", SPIE, Medical Imaging, February, 2009. Most of the work in the Section 3.2 is from this paper [83].

- S. Iftikhar, *et al.*, "An Automated Analysis for determining the effects of changes in Blood Flow Patterns around Rabbits Aortic Branches", Bioprocess UK, Conference, Brighton, UK, 2008.

Awards and Grants

- A £500 travel grant has been awarded from royal academy of engineering to participate in Single Cell Analysis Summit, 2010, San Diego, USA.

- A £500 travel grant has been awarded from IC Trust to participate in Single Cell Analysis Summit, 2010, San Diego, USA.

- A £1000 bursary has been awarded by The Charles Wallace Pakistan Trust for writing-up period for living expenses.

- Selected along with 24 other students in the UK to present my work in front of industrial people and awarded £150 bursary for presentation of a poster.

- Awarded Overseas Research Scholarship (ORS,UK), in 2007 for my PhD studies by Imperial College London, UK.

- Awarded faculty development program scholarship by Mehran university of engineering and technology, Jamshoro, in 2006.

Acknowledgements

In memory of my beloved father,

Iftikhar Rasool Soomro

whose passion for education lead me, what I am today.

Abu, you will remain in my soul, heart and mind forever!

'I know nothing.'

Sa'adia.

Abbreviations

Acronym	Abbreviation
ATP	Adenosine Tri-Phosphate
BHF	British Hearth Foundation
Ca^{+2}	Calcium
CVD	Cardiovascular Disease
CNR	Contrast to Noise Ratio
D	Downstream
DL	Downstream Left
DR	Downstream Right
EC	Endothelial Cell
ECn	Endothelial Nuclei
GFP	Green Fluorescent Protein
HDL	High Density Lipoprotein
HPAEC	Human Pulmonary Aortic Endothelial Cells
L:W Ratio	Length to Width Ratio
L	Left
LDL	Low Density Lipoprotein
NN	Nearest Neighbours
PBS	Phosphate-buffered saline
PI	Propidium Iodide
PMT	Photo-Multiplier Tube
R	Right
RBF	Radial Basis Function
SI	Shape Index
SMC	Smooth Muscle Cell
SNR	Signal to Noise Ratio
SVM	Support Vector Machines
U	Upstream
UL	Upstream Left
UR	Upstream Right

Contents

Contents

List of Figures

List of Tables

List of Algorithms

1. Introduction

1.1. Atherosclerosis and its Importance

Cardiovascular diseases (CVDs), are one of the main causes of premature death (under the age of 75) in the developing world. Heart disease alone is causing one in three deaths in both genders. According to the British Hearth Foundation (BHF) [4], CVDs were estimated to cost the UK alone £26 billion annually. In healthy arteries, blood flows around the body supplying organs with oxygenated blood, and thus maintaining a fully functional system. If damage to the inner-lining of the artery occurs, lipoprotein influx into the wall can cause plaques which lead to reduction in blood flow due to a blocked or reduced lumen of the artery, this condition is known as atherosclerosis. Formation of atherosclerosis is strongly associated with many forms of CVDs.

Atherosclerosis is a Greek word and was used by Marchand in 1904 [102], in the sense that "athero-sclerosis" means mature plaque. *Atheromatous* means lipid-rich soft tissue and *sclerotic* is hard tissue, which is collagen rich [134]. Atherosclerosis is the major cause of cardiovascular morbidity and mortality, with underlying pathological processes that may begin during childhood [45, 123, 159]. Despite having been studied since the last century, the exact causes of atherosclerotic lesion formation are not clear. The process is thought to involve lipoprotein influx into the wall, across the endothelium, and chronic inflammation. Over time, lipids accumulate in the inner wall and plaques develop, resulting in reduction or blockage of blood flow. Eventually, this condition can lead to heart attack and strokes [46, 113, 147].

1.1.1. Human and Animal Lesion Stages and Patterns

The initial stage of atherosclerosis (initial or type I lesion, which can be detected by a microscope using histological staining) starts developing with the accumulation of large quantities of oxidized low density lipoprotein molecules (LDL, a fat or cholesterol carrying plasma protein) causing formation of macrophage foam cells [68]. Then, these macrophage foam cells and lipid-laden smooth muscle cells (SMCs) form fatty-streaks, the second stage of lesion in atherosclerosis. The third type of lesion is in between early (clinically silent) and advanced stages, and it contains fatty streaks as well as collections of extracellular lipid droplets and particles that disrupt the layers of intimal SMCs [26]. A fourth lesion type is the first stage of a histologically advanced lesion, which contains dense lipid cores. Lipid cores thicken the arterial wall, and can be seen by the naked eye. The next level of lesion is usually termed atherosclerosis or atherosclerotic plaque. This new fibrous connective tissues may cause arteries to become narrower and develop fissures, haematoma, and/or thrombus. If the lumen gets smaller, the blood flow gets restricted and can cause angina (heart pain) [24]. Significant problems arise when a lesion completely blocks blood flow or develops into a more complex form, which contain hematoma, haemorrhage, and thrombotic deposits, or fibrous plaque, that ruptures [147].

Plaque rupture and endothelial erosion are the two main causes of coronary thrombosis [68]. As determined by Pasterkamp and Falk [134], the main pathological features indicating a plaque is vulnerable to rupture include atheromatous core size and consistency of the core which depends on temperature and lipid composition. The thickness, cellularity, and stiffness of the fibrous caps vary greatly as the cap thinning and reduction in collagen content increases the vulnerability of a plaque to rupture.

Following [180], four major lesion patterns can be found in human and animals around the intercostal branch ostia which show the progression of lesions with age, Figure 1.1(a). Figure 1.1(b) shows changes in lesion patterns around human intercostal branches with age. In children and in cholesterol-fed immature rabbits, downstream arrow-head lesions develop. The lateral lesion

forming at the lateral margins of ostia, and spontaneous patterns can be seen in human younger adults and cholesterol-induced rabbits. Upstream-streak occurs in adult humans and mature cholesterol-fed minipigs, progressing to the raised-lesion (which usually resembles a volcano) in old age [180].

More than 300 risk factors for CVDs have been identified in the literature [142]. These are categorised as modifiable and non-modifiable risk factors. Non-modifiable risk factors include age, gender, family history, genes, and birth weight. The major modifiable risk factors are: high LDL cholesterol, high blood pressure, smoking, low HDL cholesterol, physical inactivity, diabetes, obesity, homocysteine, and blood viscosity [142].

A question was raised by William Arid in [14], as to how these systematic disturbances get channelled into non-random, geometrically defined pathological lesions. This is usually explained by local alterations in haemodynamic forces [14]. It is therefore necessary to understand the effect of these forces on the localisation of atherosclerotic lesions.

1.2. The Effect of Haemodynamic Forces on the Localisation of Atherosclerosis

A striking feature of atherosclerosis is its non-uniform distribution within the arterial system. This is most evident in regions of branching and curvature, and has therefore been attributed to spatial variation in mechanical forces, particularly the haemodynamic wall shear stress exerted on the endothelium by the flow of blood. Throughout life, changes in arterial dimensions occur, even if a disease is present or not, due to the homeostatic system which maintains a constant perfusion throughout the body. Haemodynamic forces play an important role in vascular pathologies, particularly in relation to the localisation of atherosclerotic lesions [46].

Haemodynamics is the study of the relationship between pressure, viscous resistance to flow, and volume flow rate in the cardiovascular system [18]. The effect of haemodynamic forces may be partially explained as follows: a tangential force (blood flow) acts on the endothelial surface, resulting in a frictional
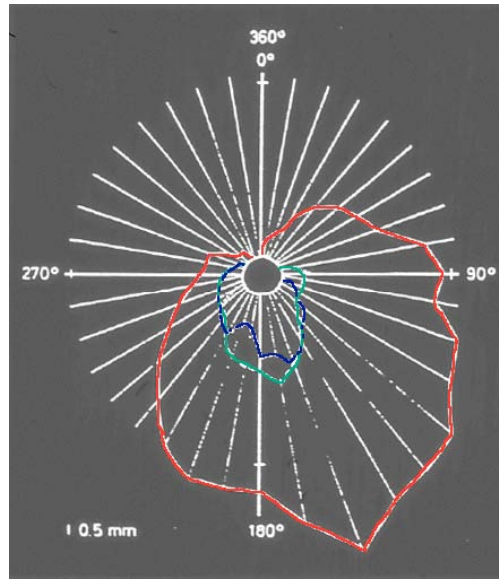
(a) Average lesion size around intercostal
branch ostia (white circle) in young humans.
Blue = Fetus, Green = Neonate, Red = Infant.
Adapted from [155].
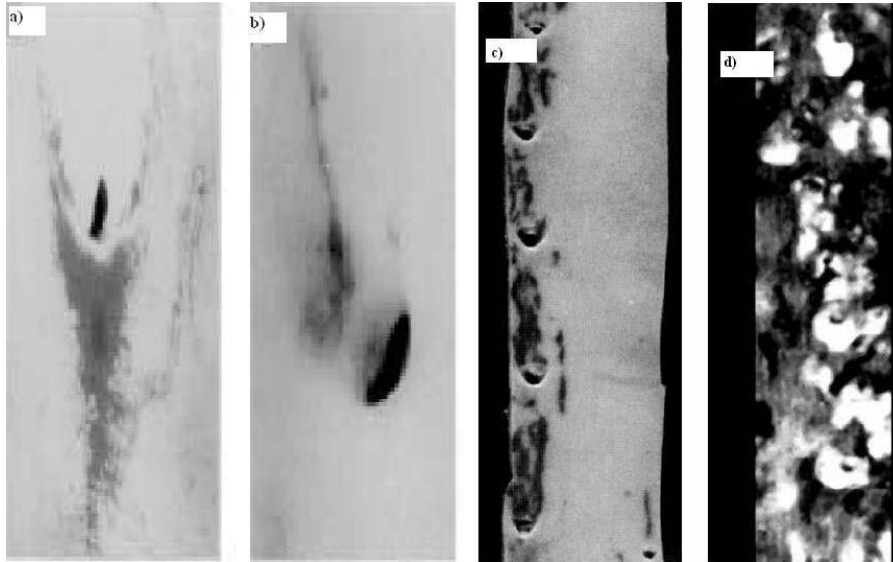
(b) The arrowhead, the lateral, upstream streak, and the volcano patterns. Adapted
from [180].

Figure 1.1.: Lesion patterns found in rabbit and human aorta. Blood flow is
from top to bottom.

force, due to viscous resistance to flow. Therefore, wherever flow occurs, shear stress exists [47]. Its magnitude can be estimated using Poiseuille's law as a frictional force per unit area, which states that shear stress is proportional to blood flow viscosity, and inversely proportional to the third power of the internal radius of the vessel [119]. Shear stress can be calculated using:

$$\tau_s = 4\nu \frac{q}{r^3} \tag{1.1}$$

Here, ν is fluid viscosity, r is the arterial radius, and q is the flow rate, and shear stress, τ_s, is usually measured in Pascals (Pa) or dyne/cm^2 (dyne/cm^2 = Pa $\times 10$). Thus, a small change in r greatly influences τ_s [47]. In different imaging modalities, different ranges of shear stress have been measured [119]. Shear stress has been shown to be the main determinant of vessel calibre and influences vessel wall remodelling [26].

Haemodynamic factors either directly influence endothelial physiology or indirectly modify local concentrations of chemicals, thereby influencing association between these molecules and their endothelial receptors [46]. Near-wall blood velocity, on which shear stress depends, cannot be accurately measured by current techniques. However, endothelial cells (ECs) form a monolayer between the blood and arterial wall [46] that is regulated by haemodynamic forces through flow mediated signal transduction [119]. Of relevance to the present study, endothelial cells and their nuclei align with the predominant flow direction and elongate in response to increased shear. Therefore, the cells can be viewed as "flow sensors" and their shape used to assess patterns of wall shear stress. For example, studies such as [38, 53], including this thesis are aimed at understanding why the pattern of disease around aortic branches changes with age [16, 25].

1.3. Role of the Vessel Wall and Endothelium in CVDs

In the cardiovascular system, the endothelium is a monolayer of cells located between blood flow and arterial wall, and has a unique response to fluid forces

helically arranged fiber-
reinforced adventitial layer

transversely isotropic fiber-
reinforced medial unit

helically arranged fiber-
reinforced intimal layer

collagen fiber s
elastic lamina externa
collagen fibril
smooth muscle cell
elastic fibril
elastic lamina interna
endothelial cell

I M A
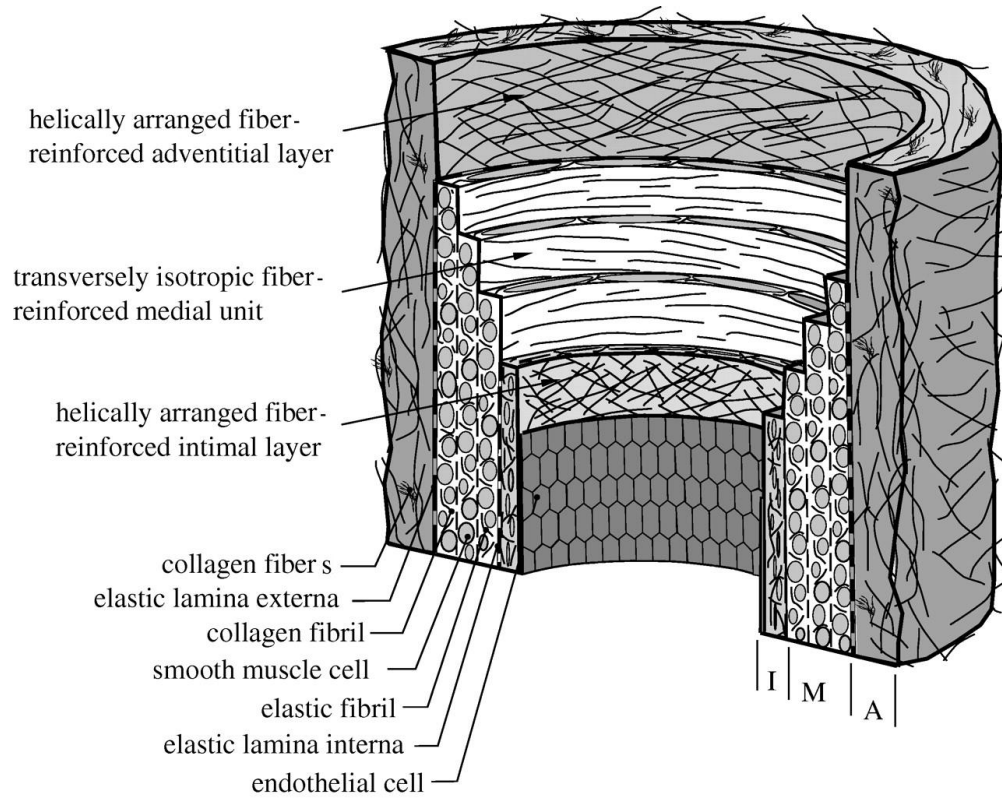
Figure 1.2.: Diagram of endothelium and vessel. Adapted from [74].

[46]. It is considered a sparse and expansive organ system, covering a total surface area of 7000 m^2, due to its vast extension, and has the ability to perform a complex array of specialized functions [14, 60]. As atherosclerotic lesions are formed predominantly in the large arteries, the structure of the vessel may play an important role in the susceptibility to lesions or resistance [24]. Except capillaries, all blood vessels are made up of three main layers, namely the intima (inner layer), media (middle layer), and adventitia (outer layer), as shown in Figure 1.2.

The intima is made up of the endothelial cells monolayer. The changes in the pathology of intima, such as deposition of fatty substances, calcium, collagen fibres, cellular waste products and fibrin, may result in the formation of athe-

rosclerotic plaque. The media or the middle layer of the artery is made up of smooth muscle cells, and elastin and collagen fibrils. The adventitia is usually made up of fibroblasts, fibrocytes and thick bundles of collagen fibrils [74].

1.4. Endothelial Cell

This and the next Section 1.5 give a brief overview on endothelial cell and its heterogeneous response, but the most comprehensive reviews in the literature are provided by W. C. Aird in [14, 15].

Of approximately 60 trillion endothelial cells in a vascular bed, every cell may be thought of as a sensing device, which receives an input from the outside environment. Each cell then undergoes phenotypic changes based on the input it sensed [14, 15].

Inputs to ECs could be both bio-chemical and bio-mechanical. Input changes in both time and space, and could be an enormous range of signals, such as growth factors, pH, oxygen, and hormones. Bio-mechanical signals are primarily shear stress and cyclical strain. ECs transduce these signals via signalling pathways that begin from the cell surface and end at the level of post-transcriptional or transcriptional modification. The phenotype or output of the cell may be a change in cell shape or calcium flux; changes in protein and/or gene expression; proliferation, migration, or cell death. ECs show heterogeneous response in their structures and functions [14], discussed in detail in Section 1.5. A schematic diagram, adapted from [35], showing mechanical forces stimulate endothelial cells through the activation of mechanosensors is shown in Figure 1.3.

1.5. ECs Heterogeneity

Endothelial cell heterogeneity occurs throughout the body. Genotypic and phenotypic differences exist between veins and arteries, between vascular beds, and even between different sites of the aortic valves [14, 60, 160]. Cell to cell variations between endothelial cells on a small spatial scale have not been studied extensively. This is rather surprising, as many researchers have measured en-
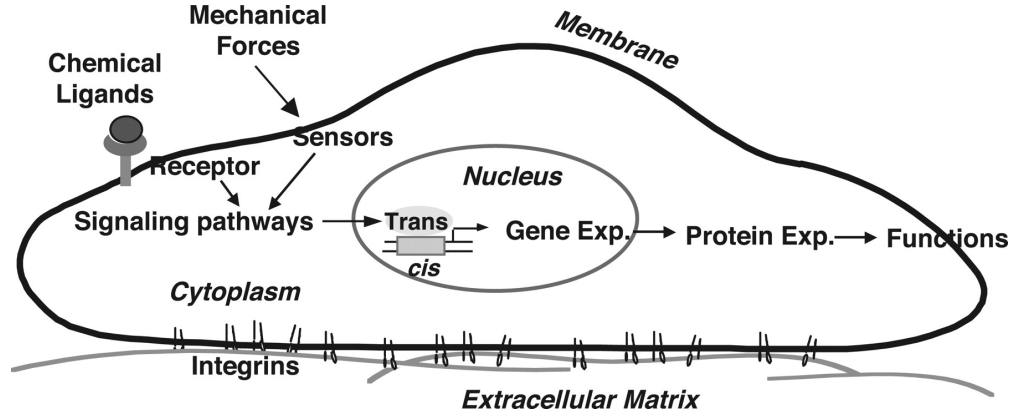
Figure 1.3.: Schematic diagram showing mechanical forces to stimulate endothelial cells through the activation of mechanosensors. Adapted from [35].

dothelial cell heterogeneity in well controlled conditions such as a parallel plate flow chamber.

Direct communication between endothelial cells occurs mainly through gap junctions, anchoring junctions and tight junctions. While tight junctions regulate permeability, anchoring junctions allow cytoskeleton proteins to interact from cell to cell and through these junctions endothelial cells may compensate for differences in stretch. Gap-junctions consist of connexin-based pores that allow direct transfer of proteins between cells. Importantly, tight junction assembly, anchoring junction formations and connexin-40 expression patterns are regulated by shear stress. This indicates that interaction between cells is shear stress dependent and therefore different shear stress levels might induce different levels of co-ordination between cells in a monolayer, like the ones, described in Section 1.7.2 of the thesis, in flow chambers.

At a cellular level, cell-cell variability has been attributed recently to stochastic gene expression [37]. These studies have been followed by more extensive studies, where besides stochastic gene expression differences in pathway capacity, differences between protein synthesis capacities were taken into account as additional explanations for cell-cell variations. Pathway capacity merely

describes the efficiency of pathways to generate and control transcription factors, while protein synthesis is due to translation of genes. It should be clear that cellular differences are considered a stochastic process and will lead to uncorrelated variation in output measures like gene expression, while cell-cell communication results in synchronisation of cells and clustering. It is likely that a combination of both processes occurs in endothelial cell monolayers.

ECs display structural and functional variations at a single point in time. From a structural perspective, endothelial cells differ in size, shape, thickness, and nuclear orientation in different locations of the vascular bed [14]. The main functions of the endothelium are the maintenance of anticoagulant properties, the physiological control of lumen diameter and vasomotor tone, blood cell trafficking and haemostatic balance. It is involved in the regulation of vascular permeability, proliferation, and the pathological consequences associated with acute inflammation, wound healing, and cardiovascular disorders such as the focal localisation of atherosclerosis [46].

1.5.1. The Role of Ca^{+2} Probes in ECs Biology

Calcium dependent pathways play a central role in the biology of endothelial and smooth muscle cells. Calcium is a major intermediary molecule in some key signalling pathways including the regulation of several nuclear transcription factors, cytoskeleton re-organisation, disruption of intercellular junction and adhesion proteins, endothelial permeability and nitric oxide synthase activity [103]. Early in-vitro experiments using a flow chamber revealed a flow dependent increase in calcium in endothelial cells, where this increase was relative to the magnitude of the applied shear stress [187, 188]. This response has been identified to begin in the caveolar area and to propagate through the cell membrane as a calcium wave. This suggests a possible localisation of one of the mechanosensory receptors in the caveolae. In addition to the intracellular wave, intercellular waves of calcium concentration have also been identified [85].

Several studies reported single-cell asynchronous calcium oscillations in response to flow [152]. The number of cells oscillating as well as the magnitude and the frequency of oscillations was found to rise as the level of shear increased.

Hemlinger, [71, 72], reported also that pulsatile flow induced considerably more single cell calcium oscillations than steady flow although the average calcium level was not changed.

Propagation of calcium waves has been proposed to occur through gap junctions, through the release of ATP and paracrine interaction. In the present research, it is argued that if ATP plays a role, the presence of transport of ATP by fluid flow would change the synchronisation of calcium waves in the presence of flow as compared to the absence of flow. However, if intracellular movement through gap junctions appeared the dominant process, then flow would not have an effect on calcium synchronisation. For calcium synchronisation, a new analysis method is developed which is discussed in detail in Chapter 6.

The above story describes the biological motivation behind the study of endothelial cells and their role in the formation of atherosclerosis lesions leading to CVDs. This forms the background to the research presented in the next chapters. It also emphasises the importance of accurate and automated analysis techniques in order to understand the disease patterns more reliably, quickly and accurately.

In order to understand spatial patterns of endothelial cells and analyse automatically, the first step is always to acquire cell data and then develop the algorithms for automated analysis. Image data will be described in Section 1.7 was provided by the Physiological Flow research group, Department of Bioengineering, Imperial College London. More specifically, the rabbit data, described in Section 1.7.1, is provided by Dr Andrew A. Bond and the *in-vitro* data, described in Section 1.7.2, is provided by Dr Nicolas Foin. The author has developed different algorithms for automated analysis based on code developed by herself and standard and modified libraries of *Matlab*, unless otherwise stated.

1.6. Different Types of Staining Techniques

The range of physical principles underlying the image contrast mechanisms of different microscopy procedures is large, and still growing. Silver nitrate staining that dates back to Ramón y Cajal and Camillo Golgi who used it to delineate dendritic structure, relies on the reduction of the nitrates to metallic

(colloidal) silver which appears black. Hematoxylin and eosin (H&E stain) is commonly used in light-microscopy for medical diagnostic purposes. Under visible light illumination, hematoxylin imparts a blue appearance to cell nuclei due to its affinity with nucleic acids; eosin, on the other hand tends to lend a pink appearance to cytoplasm. Thus, the nature of the contrast mechanism of H&E staining is an absorption-based colour shift. Other contrast systems may rely on ratios of intensities. Fluorescence-based contrast mechanisms often rely on labelling agents, or fluorophores, which absorb visible light typically emitting at a specific wavelength. For example, fluorescein angiography is a method that can be used to visualise very small blood vessels in the retina. This is used as a diagnostic technique in ophthalmology, where sodium fluorescein is injected into the patient and one or more red-free (filtered) images are acquired. In cell microscopy, it is increasingly common to use fluorescent proteins for the purpose of labelling and contrast enhancement because of their higher specificity and lower phototoxicity. For example, Green Fluorescent Protein (GFP), a protein that exhibits bright-green fluorescence when exposed to blue light, can be expressed by naturally occurring or genetically modified cellular organisms and is widely used in live-cell studies [169]. Two-photon microscopy, whereby pulsed lasers generate non-linear effects in tissue leading to fluorescence at specific wavelengths, are gaining wide usage in neuroscience as they permit confocal *in vivo* imaging of cortical tissue.

In the variety of techniques outlined above, the contrast mechanisms can be very different because the physical processes underlying the formation of the image can vary widely. For example in two-photon imaging, a photomultiplier tube (PMT) or a charge-coupled device (CCD) may be used to capture the image, whereas in GFP based imaging standard video microscope techniques may be employed. This variety of imaging systems and contrast mechanisms means that the appearance of structures can vary dramatically across studies. The structures such as statistics of object-background pixels, including correlation statistics between neighbouring pixels can indeed vary quite a lot. A very simple example might be in Micro-CT imagery where the existence of partial volume effects in cone-beam acquisition could increase the spatial correlation between pixels of an image in a structure-dependent way and the severity of

this effect would be dependent on the reconstruction algorithms. The nature of spatial correlations would be quite different in a confocal microscopy system in which out-of-plane structures could be a source of structural noise.

The variety of contrast mechanisms suggest that approaches which are based on training from hand-labelled ground-truth images might be a more efficient way of approaching segmentation problems, particularly where large data sets must be analysed.

1.7. Introduction to the Datasets used in this Study

1.7.1. *In-vivo* ECs and ECn Datasets[1]

Endothelial monolayers were stripped from the descending thoracic aortas of rabbits by a modified Häutchen procedure. In principle this involves pressing the endothelial surface of the opened aorta against a double-sided adhesive tape adhered to the surface of a microscope slide. Descending thoracic aortas of immature (6-7 weeks) and mature (> 6 months) male New Zealand White rabbits (Interfauna strain; Harlan UK) were fixed in situ at physiological pressure. The aortas were dehydrated in ethanol and cut perpendicular to their longitudinal axis to give rings containing pairs of intercostal branches. Each ring was cut longitudinally along its ventral aspect and pressed (endothelial surface down) on to a strip of double-sided adhesive tape adhered to a microscope slide. The tissue was air dried and immersed in glycerol before the intima and media were peeled away, leaving the endothelium attached to the tape. Nuclei were stained with propidium iodide (PI) after treatment with RNase. Aortas were obtained from 3 male New Zealand White rabbits (one mature and two immature) that had been perfused in situ with a silver nitrate solution[2]. A montage of images representing the area around the origin of an intercostal artery from the thoracic aorta was obtained through a Zeiss Axioplan epifluorescence microscope

---

[1]This work was done by Dr Andrew R. Bond of Department of Bioengineering, Imperial College London, London, United Kingdom.

[2]All animal procedures complied with the Animal (Scientific Procedures) Act 1986 and were approved by the Local Ethical Review Panel of the University of Reading or Imperial College London.
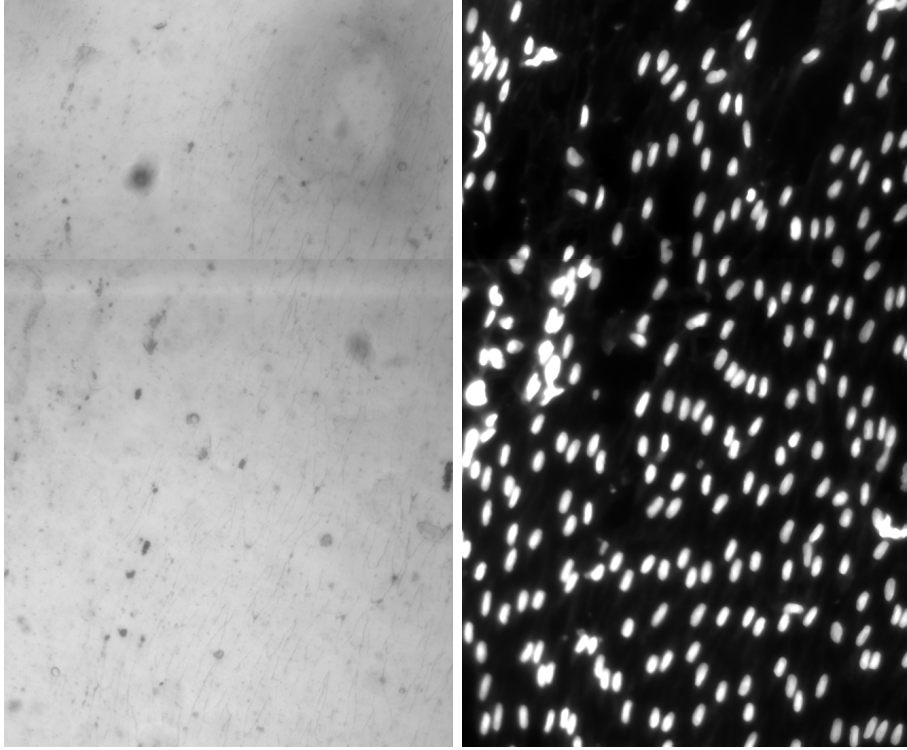
Figure 1.4.: Original sample pair of ECs and ECn images with a size of 661×1101 pixels, acquired through the method described in Section 1.7.1.

[16, 24]. Images of thoracic aorta of montage images were then subdivided as shown in Figure 1 in [16]. One pair endothelial cells (ECs) and endothelial nuclei (ECn) images of the sample images (with a size of 661×1101 pixels) taken from one of the montages is shown in Figure 1.4. Detailed information on data acquisition is given in [25].

1.7.2. *In-vitro* ECs Datasets[3]

Cell Culture

Human pulmonary artery endothelial cells (HPAEC) obtained from Clonetics (San Diego, CA) were grown on a 1% gelatin-coated tissue culture flask in M199 supplemented with 15% FBS, 2 mmol/l L-glutamine (GIBCO-BRL), 50 g/ml heparin and 30 g/ml EC growth factor [3]. The cells used in the present experiments were from passages 3 to 6. HPAEC were seeded at a 50 % cell density in IBIDI luer IV slides as described in [5]. After two days of culture the cells formed a confluent monolayer and were ready for experimentation.

Dye Loading

The calcium dye loading was performed as described by [188, 187]. In brief, the cells were washed three times with Serum free M199 and then incubated at 37 deg in 5% CO_2 for 40 min with 5 umol/L indo-1-AM [7] and 0.04 % pluoronic F127 dispersing agent (Molecular Probes, Invitrogen). The cells were then washed and incubated with complete medium for 10 min. Next, the slides were rinsed 3 times with Phosphate-buffered saline (PBS) and placed on the microscope stage. In the presence of calcium, the indo-1 dye shifts its fluorescence emission peak from 480 to 405 nm. This process is reversible and variations in calcium concentration can be retrieved by calculating the ratio of intensity between the two wavelengths.

---

[3]This work was done by Dr Nicolas Foin of Department of Bioengineering, Imperial College London, London, United Kingdom.
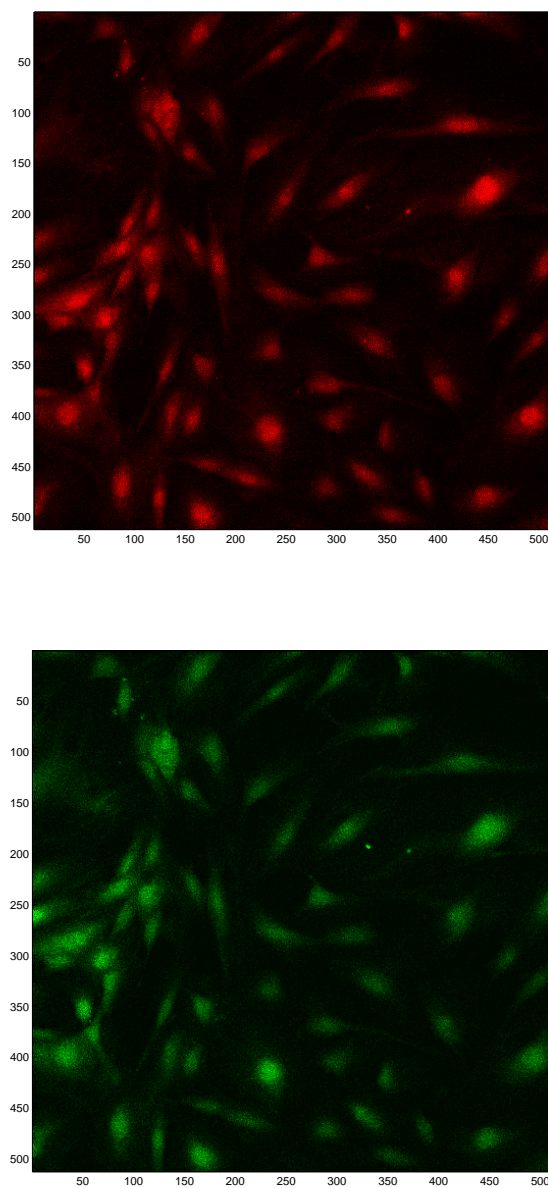
Figure 1.5.: Red and green indo-1 fluorescent images of original samples.

Shear Stress

In the flow experiment PBS was used to perfuse the chamber. A reservoir of PBS was maintained on a hot plate at 37 degrees. The flow was supplied by a peristaltic pump (Atto instruments SJ-1220, Tokyo) and the perfusion rate was determined to produce the shear stress required by the slide manufacturer specification [5].

Microscopy

Fluorescent images from Indo-1 were obtained with a confocal laser microscope equipped with an UV argon ion laser (MRC-1000 UV, Bio-Rad) as previously described in [187]. A laser with a 351-nm wavelength excited cells through a × 40 objective. The light was separated into 405 and 480nm wavelengths by a beam splitter before reaching the photomultipliers. One sample pair of red and green fluorescent images is presented in Figure 1.5.

1.8. Related Issues

In this thesis, we will address the problems related to automated analysis of *in-vivo* cell data. Among these, one was the detection of weak cell boundaries with low-contrast and low signal-noise ratios between cell boundaries and the background. Some of this thesis work was aimed at developing low-level vision algorithms to address the segmentation problem. In the data given, some cell boundaries were so thin, weak and noisy that they were hard to discern by eye. Weak cell boundaries can be described as those which have very low contrast or signal to noise ratio (SNR and CNR). A demonstration of the SNR and CNR is given in Chapter 3.

In the ECn data, there was significant intensity variation among nuclei. The problems which needed to be addressed in analysing these images of nuclei include proper selection of the intensity threshold. Another problem was that some nuclei overlapped with multiple other nuclei. A solution for overlapped nuclei was to discard these and work only on isolated cells and nuclei in the images. This was sufficient to enable further analysis and to supply data in the

quantities needed.

When performing shape analysis, an interesting challenge that was faced was the question of the numbers of identifiable patterns, or modes, of cell shapes.

In *in-vitro* data analysis, the biggest issue was to measure the cell movement and change in intensity over time in each cell. There were some cells which overlapped in the given data. Those were also discarded to do single-cell analysis.

1.8.1. Challenges and the Problem of Analysing Large Datasets

With the introduction of computerised systems in hospitals of the developing world for diagnosis and pathology, an enormous amount of medical image data is produced each year. With the increase in the size and complexity of the datasets, it is important and necessary to develop robust techniques to analyse, evaluate and interpret the information contained in these rich datasets. Their large size poses different challenges from the methods developed for small or moderate sized datasets.

Automated analysis is needed for modeling and quantitative analysis of these large data sets, simulations, and sometimes real time interventions, as manual assessment of medical images is a tedious, error prone task that often requires an expert for the analysis and to give an opinion.

1.9. Project and Thesis Aims

The main contribution of this thesis is to develop algorithms to analyse the spatial patterns of endothelial cells, and the functional behaviour of endothelial cells under different flow conditions, in order to understand the patterns or formation of atherosclerotic lesions.

During the course of this research, two different projects were tackled regarding understanding endothelial cell behaviuor and analysis. The two main sub-projects are:

First

To build an automated analysis system for microscopy cell images to gather information on large numbers of cells automatically. Also, to study changes with age in the pattern of wall-shear stress by examining shear-dependent morphological features of ECs and ECn. In order to achieve this, several steps are needed, including segmentation of the cell nuclei and boundaries, associating cells and their respective nuclei, and building a shape analysis technique to identify patterns in cell shape in different animals or within different regions of the blood vessel.

Second

To analyse *in-vitro* data acquired from video sequences of fluorescent microscopy of cells under flow conditions. The main objective here is to draw inferences about pattern formation of cell behaviour, and the functional behaviour of cells during flow conditions. Cell-cell heterogeneity, as discussed in Section 1.5, at cellular-level has not been studied extensively, though it may play a role in progression of disease. In order to analyse cell behaviour, single-cell probes are needed, and in this thesis a Ca^{2+} probe, as discussed in Section 1.5.1, responses to the flow are used for the analysis to get these answers: How do cells interact with each other (synchronicity)? Are there groups among cells? Is there any sentinel cells, which initiates internal cell signalling and communication processes?

1.10. Thesis Structure

Chapter 2 reviews the image segmentation techniques that are relevant to the cells and nuclei segmentation. Some classification techniques are also presented which are applied to cell and nuclear data.

Chapter 3 presents different methods explored in this research for segmentation of EC nuclei (ECn) and EC boundaries (ECs). As the detection of the nucleus was an easier task than the cell itself, due to higher contrast and signal to noise ratios, different proposed segmentation algorithms are presented for ECn and more specifically for ECs boundaries.

Chapter 4 gives a comparative analysis between stress effected morphological parameters of EC boundaries and EC nuclei in order to understand the age-related changes in patterns in rabbit data.

In Chapter 5, shape analysis techniques developed for EC boundaries and nuclei are presented to detect and classify morphological and variational modes of the shapes in EC cells and nuclei.

Chapter 6 describes the EC behavioural analysis in controlled flow and some techniques developed during this thesis work to detect the heterogeneity among cells, more specifically the formation of intricate patterns in Ca^{2+} probes among themselves and their neighbourhood.

Chapter 7 presents the main findings, outcomes achieved and conclusions made from this research, with a critical overview of the advantages and limitation of the methods developed. Some suggestions for improvements upon the current work are provided in the Future Work section.

2. Review of Cell Segmentation and Classification Techniques

2.1. Introduction

Automatic analysis of cells and their nuclei may require segmentation, registration, labelling, and classification tasks to be performed [158]. First, segmentation is performed to isolate the object of interest (cells and nuclei in this case). The purpose of labelling and classification is to extract useful information about isolated objects and assign them to different classes. This Chapter gives a brief overview of relevant techniques.

2.2. Cell Segmentation

Segmentation techniques for automated analysis can be broadly categorised into **feature based**, **edge-based**, **region-based**, and **model-based** techniques. In this section, a brief survey of the latest cell and nuclei detection and segmentation techniques is given. For detailed image segmentation techniques, the author recommends [30, 58, 59, 69, 81, 117, 138, 141], as review papers.

Segmentation of cell regions is an important step in automated analysis of medical images. It is always difficult for a machine to partition images into regions without prior information. In broader terms, there is no universal method or general algorithm to segment an image for all data types and contrast mechanisms. One of the reasons for this is that a 2-D image can be represented in an infinite number of possible ways. Another reason, as described by Palvidis in [77], is that an image segmentation problem involves psychophysical perception of an image and might not be susceptible to any analytical treat-

ment. Algorithms often use heuristics and *a priori* knowledge about the image to be segmented.

Segmentation involves partitioning an image into its constituent parts or regions so that the union of these regions recreates the complete image, $f(x, y)$, as described below in Equation (2.1), where R_i is the ith region:

$$f(x, y) = \bigcup_{i=1:n} R_i \qquad (2.1)$$

In general, the regions in which the image is to be segmented must have some common criteria or features such as grey level, variance, shape etc. There are some conditions for the above equation to be satisfied as described in [59]: 1) the segmentation step should not terminate until all pixels in an image have been processed, 2) each region must be contiguous, and last but not least, 3) regions should meet some similarity criterion.

There are two main reasons for the complexity of the image segmentation task in biological and medical image data. Firstly, biological structures possess a degree of natural variability that is not easily captured by simple models of deformation: the natural modes of variation are surprisingly broad, even if one excludes pathological conditions. In contrast, in man-made object recognition projective models of view transformation have proven very powerful when used either for view invariance, or for inferring structure from tilt. On the other hand, once biological structure labelling has been achieved, and pixel locations along a two or three dimensional object have been determined, modes of variation have successfully been analysed to study some aspects of biological variability (see, for example, [101]). Such studies have also had applications in face processing, in film post-production and also in forensic analysis, where the effects of ageing, for example, may be predicted through the principal components of variation over time. However, this author is not aware of widespread application in studying biological shapes acquired through microscopy.

The second reason pertains to the variety of contrast mechanisms found in biological imaging. In tissue or biological fluid microscopy, the contrast of a sample can be dramatically altered by the choice of staining. Creating algorithms that are robust to these changes requires tolerance not only to scale and

slice orientation changes (for tissue), but also to the property of texture, which is highly variable in biological data, even within a given staining.

Much of the research in cell image analysis is done on the nuclei. There are mainly two reasons for this: (a) most of the diagnostic information is associated with the nucleus [195], and (b) segmentation of the nucleus is much easier than of the cell itself. The contrast between cytoplasm and extracellular space can become very weak, even undetectable, such as, in case of this study, for finding the ECs boundaries. Due to complexity of the images, even detection of the nucleus is not always trivial [196].

Segmentation techniques in general, and cell segmentation techniques in particular, are often divided into the following types which are reviewed in detail in the next sections:

1. Feature-based techniques

2. Edge-based techniques

3. Region-based techniques and

4. Model-based techniques

2.2.1. Feature-based Techniques

Thresholding

Researchers have developed various techniques to overcome difficulties of cell and nucleus segmentation. The simplest method for isolation of cells and nuclei from tissue images is thresholding [138]. Feature-based segmentation is a special multi-dimensional case of thresholding. The basic idea of feature-space segmentation techniques is to partition a $\mathbf{d}$-dimensional feature space, where $\mathbf{d}$ is the number of dimensions of feature space, computed for each pixel to be segmented. A feature vector is a $\mathbf{d}$-dimensional vector of features, usually statistical measurements that describe some properties of each pixel. The number of feature vectors is usually equal to the number of pixels in the image. Each pixel in the original image is mapped to a location in feature space based on its local properties. Once the feature space has been generated from all pixels

in the image, it is then partitioned. Then, the feature space is mapped back to the original image space to produce the segmented image.

Even though, there is no set of rules defined for selecting the features for any application [36, 135] in general each feature should be:

1. Unique and parsimonious (implies little or no redundancy),

2. Scaling, rotation and translation invariant,

3. A complete, congruent, and compact set.

Some other properties could be:

1. Ability to reconstruct the image from its descriptors,

2. Physical interpretation: features should relate to some physical properties,

3. Features should be gathered automatically to avoid or reduce human bias or error.

Figure 2.1 (a) shows an example of a partitioned 2-$\mathbf{d}$ feature space with two apparent classes. A more detailed description on feature selection is given in [129]. Once the features are selected, the next step is feature normalisation. In a normalisation stage, these features are transformed in such a way that all should become dimensionless. The simplistic step for normalisation is given in Equation (2.2):

$$\mathfrak{N}(i, \mathbf{d}) = \frac{\mathfrak{F}(i, \mathbf{d}) - \mu_{\mathbf{d}}}{\sigma_{\mathbf{d}}}, \quad \forall \mathbf{d}, i \tag{2.2}$$

Where $\mu_{\mathbf{d}}$ and $\sigma_{\mathbf{d}}$ are the population mean and standard deviation of the $\mathbf{d}^{th}$ feature and i^{th} element of input space, $\mathfrak{F}$ is the feature space and $\mathfrak{N}$ is the normalised feature space with a zero mean and a unit variance. This reduces the arbitrary range of some dimensions of the feature vector. Feature domain segmentation can be either supervised or unsupervised. The detailed review of these techniques is presented in Section 2.4.

23

Thresholding is a 1-**d** feature-based (that feature usually being greyscale) segmentation technique. For greyscale images, thresholding is a way to distinguish pixels of lower intensity from higher intensity values. The thresholding operation can be expressed as:

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq t, \quad \forall(x,y) \\ 0 & otherwise \end{cases} \tag{2.3}$$

where $g(x,y)$ represents a binary image of the greyscale image $f(x,y)$ and t is the threshold value.

In thresholding, the assumption is made that object and background pixels in the image can be distinguished by their grey-level values. For example, in an image of a tissue section, cells can be segmented from their background with a grey-level threshold, t, by analyzing dominant values such as the gray level of the cells and background intensities carefully. Figure 2.1 (b) shows the histogram of a scalar image that possesses three apparent classes, corresponding to the three modes. Thresholding techniques are relatively useful in microscopy but can be improved [131]; the main problem is to determine an appropriate threshold. In practice, it is impossible to find a single threshold level to suit a whole tissue section image, therefore additional information is needed. The problem can be avoided by the use of local thresholds, as described by Niblack [128]. In early stage of this thesis, a modified version of Niblack's method was presented in [83] to automatic cell boundary detection. This method, described in detailed in Section 3.2, is limited by the problem of accurately finding a binary map representing the locations of image cell boundaries. In [148], Sahoo *et al.* divided thresholding techniques into point dependent global, region dependent global, local, and multi thresholding. In [178], an image analysis program was applied to count individual cells in tumours labelled for proliferation markers, by an algorithm which selected thresholds automatically. Similarly, the grading of bladder tumours is presented in [124], based on local thresholding. These methods were developed for the segmentation of nuclei and background. In the tissue sections, automated detection of cells can be found in [116, 133]. A watershed algorithm is applied in [174] to determine the contours of cells. In [27], cytoplasm was separated from background in cervical
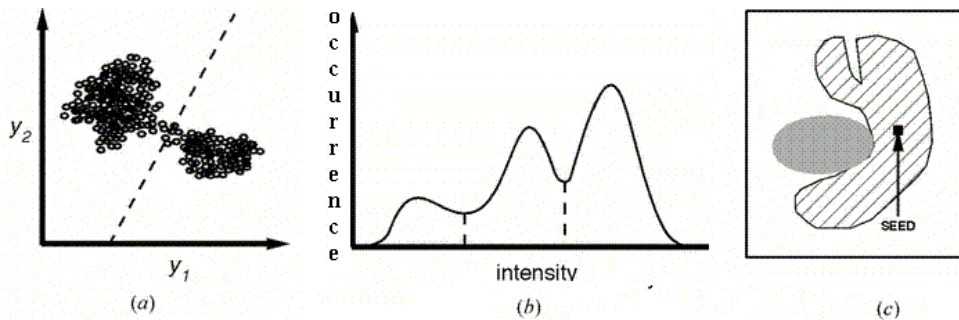
Figure 2.1.: (a). Feature space method (b). histogram method and (c). region growing. Adapted from [138].

cell images by varying the threshold using the stability of the perimeter of the cell. With the determination of a threshold, cytoplasm and nucleus were clustered into cytoplasm, folded cytoplasm and nucleus. Surveys on thresholding techniques are presented in [148, 150].

Mathematical morphology approaches combined with other segmentation techniques also have been widely used to isolate cells and nuclei, and find morphological features, such as granulometric moments in [162]. Another study [163], applied morphological operations followed by extraction of nucleocytoplasmic ratio and hyperchromasia features to segment the nuclei.

2.2.2. Edge-based Techniques

Edge detection is a low-level technique applied to images in order to detect discontinuities or sharp changes in image brightness. Segmentation techniques based on edges or gradients use the discontinuity of image intensities or texture at the boundary between objects, but these methods are sensitive to noise and do not work well on broken contours [95]. The performance of basic edge detectors is reviewed in [167].

In [91], Kaspersen *et al.* introduce a continuous wavelet transform method for detection of edges in ultrasound images. An example of cell detection using edges may be found in [196]. Edges can also be detected in histological images by applying the Laplacian operator, and Laplacian of Gaussian (LoG) operators, such as in [12, 126]. In [164], a variation of the Hough transform for

2. Review of Cell Segmentation and Classification Techniques

ellipses has been proposed for the determination of the location of the nuclei. The result obtained with edge detection is usually a set of edges or sometimes broken edges plus artefacts or noise. Additional processing is usually needed to reconstruct the edges or borders of the regions. The Canny edge detector [29] along with morphological operations is another useful method, which I have also employed for segmentation of ECn, to be discussed later in Chapter 3. Various methods have been developed to perform edge linking, such as sequential edge linking combined with threshold in [179] to extract the contour of a region of interest in a mammogram while in [153] an application is presented to fill in the gaps between edge segments.

2.2.3. Region-based Techniques

Region-based algorithms use region growing, splitting and merging of regions with different homogeneous intensities or properties being examined to partition the areas of interest. These algorithms are more time consuming and work best on images which are illuminated evenly [182]. Further applications of region growing can be found in [107, 170]. These algorithms require a seed point for the start which is usually entered manually, as shown in Figure 2.1 (c), in which region growing has been used to isolate one of the structures. These algorithms are mostly used with other segmentation techniques.

Low-level segmentation techniques discussed so far mostly require a considerable amount of user guidance, and automating these techniques is very difficult when the shapes are complex or variable. Noise and other image artefacts can result in incorrect regions and boundaries using these methods [58, 59, 138]. Deformable model based segmentation, discussed in the next section, can overcome many of these limitations.

2.2.4. Deformable Models

Deformable models, also known as active contour models or snakes, are introduced in [92]. Snakes, for example are curves defined within an image domain that can move under the influence of forces derived from the image data [184]. Snakes have been used in many medical imaging applications such as

26
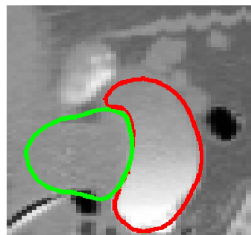
Figure 2.2.: Segmentation of bladder and prostate using deformable model. Adapted from [41].

[31, 33, 54, 88, 106]. An application of snakes for the detection of the boundaries of nerve cells can be found in [185]. Other examples of cell detection by snakes can be found in [112, 136, 165, 186]. Figure 2.2 shows the segmentation of bladder and prostate using deformable model [41]. The difficulties encountered with snakes, especially in irregular shapes or noisy images are: initialization of the snake, which must begin close to the candidate boundary so that it can converge correctly and requires the initial shape of the contour to be defined, difficulty in dealing with concave boundaries, and susceptibility to noise which can cause errors when the boundaries are not smooth. Due to the iterative nature of active contour methods, they may be slow and computationally expensive. Some of the researchers have tried to address the snake initialisation problem, such as in [104, 143]. A candidate shape model developed for fitting ellipses on to cell boundaries will be presented in Section 3.5. This model tries to find the boundaries of cells with an elliptical hypothetical orientation field and the candidate cell orientation field derived from the image. A critique of this model *vis-a-vis* other approaches will be given in Chapter 3.

2.2.5. Statistical Models

To overcome difficulties encountered by the above discussed segmentation techniques, another possible way is to use statistical models. An approach using 2-D Markov random field in polar coordinates with the use of a statistical Bayesian framework for local smoothness of image data is proposed in [34] to automatically detect the boundary of the left ventricle from echocardiographic image. H. Knutsson *et al.* in [97], introduced a new method of segmentation and regis-

tration with the use of a Morphon. The Morphon approach allows an intuitive interface for specifying prior information. They show its performance on a 2-D ultrasound heart image sequence with the aim of segmenting the heart wall. Wu *et al.* [181] selected a different approach for cell segmentation, by initializing a parametric image to approximate the cell image. Statistical models have been proposed to track the cell and boundary information such as in [195].

2.2.6. Other Methods

Several other complex and computationally intensive schemes have been developed to extract various types of boundary information, such as wavelet transforms [91]. Methods for cell nuclei microscopy data analysis include maximum likelihood methods of radial search [126]. A modified version of this technique was initially tried on our ECn and ECs images, but was found to be unsuitable, for reasons that will be discussed in Chapter 3.

2.3. Segmentation Performance or Validation

Despite the fact that much research is done in this field, it is very difficult to compare any two segmentation algorithms, unless they are implemented on the same data, as many of these programs are optimised for a specific problem and for a specific imaging modality. Validation is necessary to quantify the performance of a segmentation method. The most straightforward approach to validate a segmentation technique is to compare the automated segmentation with a manually obtained *ground-truth* [183]. This approach does not guarantee a perfect truth model, due to human observer's flawed performance. If a *ground-truth* is available, the quantification of accuracy or precision is application-dependent, it may be based on region information or number of correctly classified or misclassified pixels [138]. Segmentation software for medical images is usually evaluated by comparing its results with a segmentation done by experts. An algorithm can also be evaluated by analysis of synthetic or physical phantom images [192]. There are several databases available for different imaging modalities with their respective manually identified (label-

led) counterparts, which can be used to evaluate or check the performance of segmentation algorithms. These include the MNI [1], SBIA [2], IBSR [6] and STARE [76] databases. Some reviews on the evaluation of image segmentation techniques are presented in [191, 192, 193].

2.4. Classification of Cells in Image Analysis

Due to the complexity of nature, classification of objects in medical images is as difficult as the segmentation, described earlier in this chapter. Solutions may be sought using different approaches of artificial neural networks and machine learning algorithms. These techniques can also be applied for detecting or segmenting the objects of interest. We will present in Chapter 3, a method of segmenting cell boundaries from the background implemented with a machine learning algorithm, Support Vector Machines (SVMs), discussed in detail later in this section.

Classification is a decision making process, and refers to the task of assigning a number of classes to a **d**-dimensional feature space. Let **f**, be a feature vector in a feature space, $\mathfrak{F}$, then the objective is to assign $\mathfrak{F}$ into one of n classes, represented as $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$. Mathematically, it can be formulated as:

$$\mathfrak{F} \overset{\Phi}{\mapsto} \Theta, \quad \mathfrak{F} \in \mathbb{R}^d$$

This mapping, Φ, is usually done with a mathematical function having some adjustable parameters. These are typically determined using information provided by a set of data for which the expected mapping output is known. Classification techniques can be divided into supervised and unsupervised learning.

2.4.1. Supervised Classification

The type of classification in which a partitioning of a feature space generated from an image is done with its *ground-truth* or labels or known classes, is also known as *supervised classification* in machine learning. Classifiers are called supervised methods because they require training data before they can attempt automatic classification of new unseen data. There are a number of ways in which training data can be applied in classifiers.
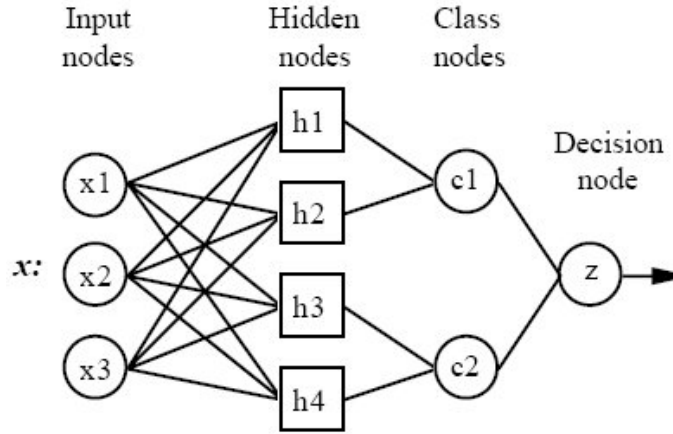
Figure 2.3.: General ANN model. Adapted from [8].

Artificial Neural Networks

Artificial neural networks (ANNs), are generally non-linear systems and are motivated by biological neural networks [86]. Neurons are their basic building blocks. A neuron takes a number of inputs with a weight assigned to each input to generate a single weighted output. A general ANN model, adapted from [8], is presented in Figure 2.3. One of the main advantages of ANNs over other classification techniques is that they can approximate any function through a training process [158]. ANNs have been used extensively over the past four decades for both classification and clustering [86].

Some ANN applications have been reported in cytological and histological analysis. For example, ANNs have been used for the segmentation of biological tissue sections for tumour grading in histological images in [89]. Different approaches for counting cells in tissue sections using ANNs is given in [156, 82]. ANNs can be used as a classifier such as in [67], as well as a clustering method such as [20, 144, 145] or as a deformable model [173].

Support Vector Machines (SVMs)

Conventional statistical learning methods tend to find functions to map high dimensional feature vectors to their respective classes, therefore, empirical risk

minimisation (ERM) of training examples does not imply good generalisation to novel test data. However, SVM learning is based on structural risk minimisation (SRM), which addresses the problem of ERM, by minimising a generalisation bound error (a sum of learning error rate made on novel test or unseen data).

The idea of SVMs, a class of kernel-based machine learning algorithms, being used for classification and regression was proposed by Vapnik *et al.* in [171]. For generalised error, Vapnik-Chervonenkis (VC) dimension has been mostly used as a quantitative measure. SRM can select minimum error on the unseen data by minimising empirical error of a function on training examples and selecting a function with a low VC dimension. This is the reason SVM has a potential to perform well on unseen data; and has outperformed many competing methods [51].

The idea of VC dimension can be represented by using hyperplane classifiers, so the main objective of SVM classifiers is to determine a decision boundary or a *hyperplane* between classes of data observed in a high-dimensional input space i.e. multivariate data. As shown in Figure 2.4, margin m is the distance from the hyperplane to the closest data point for two classes of input data. The input space or data for the SVM to classify could be either linearly or non-linearly separable. A brief description of SVMs is presented here, and the most of the following formulations are adapted from [151], unless otherwise cited whenever needed.

Separable Data

The main component of an SVM is known as the *maximal margin classifier*, and it works only for data which are linearly separable in the feature space.

Let, $\left[(x^1, y^1), (x^2, y^2), \ldots, (x^l, y^l)\right] \subseteq (X \times Y)$, represents a given l training examples set. $X \subseteq \mathbb{R}^d$ denote the vectors lying in the input space, assumed Euclidean, $Y \subseteq \{+1, -1\}$ is the label data for two classes, x^l are the training examples and y^l are the labels.

Then, the goal of the SVM is to define a functional margin of the training set with respect to a *maximal margin hyperplane* to separate all the data instances

31
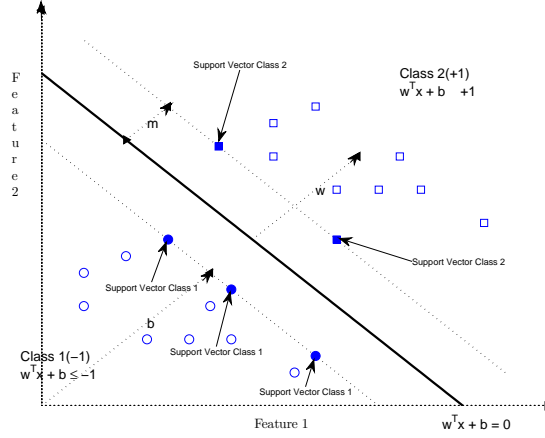
Figure 2.4.: A 2-class SVM maximal margin hyperplane in a 2-D feature space, with its support vectors filled o and □. Adapted from [42].

[151]:

$$
\begin{aligned}
\mathbf{w}^T \cdot x^i + b &\geq +1 \quad \forall y^i = +1 \\
\mathbf{w}^T \cdot x^i + b &\leq -1 \quad \forall y^i = -1, \quad \forall i, \quad i = 1, \cdots, l
\end{aligned}
\tag{2.4}
$$

where $\mathbf{w}$ is the hyperplane coefficient vector and b is the bias term [98]. Mathematically, this hyperplane can be found by minimising the cost function:

$$
\frac{1}{2}\left(\mathbf{w}^T \cdot \mathbf{w}\right) = \frac{||\mathbf{w}||^2}{2}
\tag{2.5}
$$

subject to Equation (2.4), and therefore to maximise m, the distance between 2 classes.

Non-separable Data

Of course, data may not always be linearly separable as in Figure 2.4. To account for this, non-negative slack variables, denoted by ξ^i, can be introduced, leading to

32

$$\mathbf{w}^T \cdot x^i + b \geq +1 - \xi^i, \quad \forall y^i = +1$$
$$\mathbf{w}^T \cdot x^i + b \leq -1 + \xi^i, \quad \forall y^i = -1$$
$$\xi^i \quad \geq \quad 0, \qquad \forall i \quad i = 1, 2, \ldots, l \qquad (2.6)$$

where, the ξ^i must be greater than 0 and $\Sigma_i \xi^i$ is an upper bound on the training errors. Hence, the objective function to be minimised becomes

$$\left(\frac{||\mathbf{w}||^2}{2} + C \left(\Sigma_i \xi^i \right)^\alpha \right), \quad \alpha > 0, \qquad (2.7)$$

The cost function in Equation (2.7) constitutes the so-called *structural risk*, which balances the *empirical risk* (i.e., the training errors). C, usually chosen by the user, is the penalty to errors and controls trade-off [51].

Kernel Technique

In place of using non-linear curves, one may map the data into higher dimensional feature space by $\Phi(x) : \mathbb{R}^d \mapsto \mathbb{H}$, and then find a linear solution in $\mathbb{H}$. That is, if $\mathbb{H}$ is a Hilbert space, there must be a kernel function in the original space, $\mathbb{R}^d$, that describes the inner product in $\mathbb{H}$ [171]. This is the basic idea of kernel functions, $K(x, z) = \Phi^T(x) \cdot \Phi(z)$:

$$K(S^i, x^i) = \Phi^T(S^i) \cdot \Phi(x^i) \qquad (2.8)$$

There are several transformation functions, some typically available kernels are given in Table (2.1), here, γ determines the area of influence S^i has over the data space, S is the number of support vectors, δ, and p are polynomial co-efficient. A more detailed mathematical description of SVMs is given in [151].

Some applications of the SVMs as a classifier in microscopy images which have been reported previously are presented here. For example, in [110], Genovesio *et al.* used SVMs at pixel level to segment and detect tuberculosis. In

Table 2.1.: List of different kernels typically applied in SVMs.

Kernel	$K(x,z) = \Phi^T(x) \cdot \Phi(z)$
Radial Basis Functions(RBF)	$exp(-(x^i - S^i)^2 \cdot (-\gamma))$
Sigmoid	$\tanh(\gamma(x^i \cdot S^i) + \delta)$
Polynomial	$(\gamma(x^i \cdot S^i) + \delta)^p$
Linear	$(x^i \cdot S^i)$

[63], the grading of tumours by SVMs was reported to perform well on both low and high-grade tumours, able to classify at the rate of 87.5% and 83.8% respectively. In [109], Lee *et al.* segmented different areas of brain tumour with the help of 6 classifiers, among them one being was the SVM. Yong Fan *et al.* [52] showed that brain images of normal subjects and schizophrenic patients could be discriminated with high accuracy (91.8%) with the help of the SVM. Issam *et al.* [51] applied an SVM approach to the automatic detection of microcalcification clusters in digital mammograms. Automatic classification of liver disease using SVMs was presented in [108], for differentiating between hepatoma, cysts and cavernous haemangioma and also between normal tissues, in CT images. Kalatzis, *et al.* in [90] have used SVM as a classifier to distinguish between diabetic and non-diabetic brain image samples with 99.1% reported accuracy. For automated characterisation of skin lesion images (malignant melanoma and dysplastic naevus), Ilias *et al.* in [84] reported 94.1 % performance accuracy of SVM. Hua *et al.* [79], applied an SVM to predict the subcellular localization of proteins from their amino acid compositions. Similar to this study, in [80], Murphy *et al.* successfully classified sub-cellular patterns in multi-cell florescence images. In [168], Tscherepanow *et al.* use an SVM as a classification tool to distinguish cells from non-cells automatically. In [100], it was reported that SVM and a Random Forest Method achieved the best accuracy in cell classification and feature reduction. In [70], N. Harder *et al.* trained SVM for the classification of different phases of nuclei. Wang *et al.* [177] proposed a new online Support Vector classifier to classify cell nuclei into different phases.

Pierna [139] proposed a system which analysed many samples using either partial least squares (PLS), an ANN or SVM classification models in an effort to improve the enforcement of legislation banning the use of meat and bone

meal (MBM) in compound animal feeds. The methodology combined imaging spectroscopy and chemometrics. While all three chemometric classification algorithms performed admirably in analysing the training data, and when analysing the data of spectral types which were represented within the training data set (vegetal and MBM particles), the ANN and SVM models showed superiority over PLS in generalisation ability when unmodelled data were encountered (i.e. background pixels). Although in some applications, such as that of Pierna, the performance of the SVM is not necessarily better than a well-trained ANN, it is considered that the dangers of overtraining a learning system are reduced in SVMs, partly because of the nature of the *maximal margin* as a criteria for learning to classify patterns. A comparison of segmentation results for vascular EC boundaries between different thresholding techniques, with the results obtained from a model developed for using SVMs, is presented in Chapter 3.

2.4.2. Unsupervised Classification

Cluster analysis, a method of unsupervised learning, separates the feature space into subsets (clusters), based on some mutual similarity of subset elements. Similar objects should reside in the same clusters, and different objects in different clusters. Various approaches to data clustering can be taken, but there are two main groups of methods; hierarchical and partitional clustering [86].

Hierarchical methods

Hierarchical clustering is an important category of clustering methods; the results of hierarchical clustering may be represented in a tree structure called a dendrogram. The root of the tree forms a single large grouping of data, while the leaves represent individual observations or proximities, Figure 2.5.

There are two main approaches for creating hierarchical clustering: divisive and/or agglomerative. In divisive hierarchical clustering, clustering starts from the root and recursively splits the clusters till the tree becomes all leaves; agglomerative clustering starts from the leaves and successively merges clusters together. Merging or division is performed according to some criterion or parameter, e.g. the distance or similarity between feature vectors of the objects.
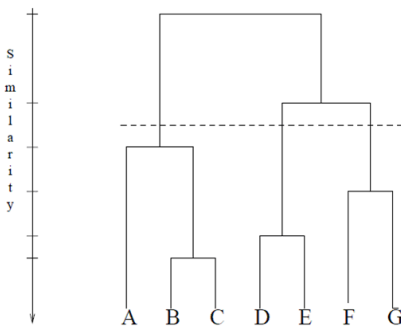
Figure 2.5.: A dendrogram obtained using a single-link agglomerative clustering algorithm, — represents the similarity distance chosen to find the number of clusters in the given data. Adapted from [86].

Partitional clustering methods

Partitional clustering methods sequentially assign each pattern to one cluster, and can be divided into parametric and non-parametric. Non-parametric clustering analysis is a most simple, practically useful approach to cluster analysis. The most common method in this category is the K-means cluster analysis method given by MacQueen [118].

Parametric clustering analysis techniques are based on class-conditional probability distributions and require a distribution parameter estimation such as the minimum error criterion [158]. The computational complexity of probability density estimation methods varies according to the amount of prior information. This is used to describe the shape of the density functions, and hence find their unknown parameters. This is why this class of learning methods is also known as parameter learning. More detailed descriptions of cluster analysis methods and algorithms are given in [56, 86, 99, 158].

Clustering techniques were applied for cell nuclei segmentation by Walker *et al.* in [176]. K. S. Fu *et al.* [114] proposed to segment blood cell neutrophil images using iterative segmentations. With some basic assumptions on the image, an initial segmentation was obtained, which was then used to determine some critical information from the image. This information then determined the number of clusters. Aggarwal *et al.* [13] used a combination of threshol-

36

ding and clustering techniques to segment cervical smear images. Thresholding at a pre-set grey level was used to extract the nucleus. 233 images were tested, and the nuclei were successfully extracted with a reported accuracy of 87.6%, and 88.1% success rate for isolating cytoplasm. In [127, 172], Murphy *et al.* describe a systematic approach for interpreting subcellular patterns in light microscopic images, using a combination of supervised and unsupervised clustering methods, with the help of a rich set of sub-cellular location features (SLF). In [30], Di Cataldo *et al.* presented an automated analysis system based on two methods for segmentation of cancerous immunohistochemical tissue images; one method allowed detection of cancerous areas with a reported accuracy of 91 % and another achieved automated segmentation of the nuclear membrane of cancerous cells with a reported accuracy of 83%.

2.4.3. Other Clustering Algorithms

A review of the many other clustering algorithms that have been developed is beyond the scope of this thesis, but these include: squared-error clustering, graph-theoretic clustering, nearest neighbour clustering, and fuzzy clustering. These algorithms are reviewed in detail elsewhere [86, 99].

2.4.4. Clustering Validity

Cluster validity analysis is the assessment of the output, and is a measure of goodness of a clustering algorithm. Often, this analysis uses a specific criterion of optimality, and this criterion is usually based on subjectivity, hence there are few *gold-standards* in clustering [86]. The desired number of clusters is generally specified by the user to start the clustering process, and the number of clusters chosen is always based on some assumptions made by the user about the data. It is still unclear how to find the correct number of clusters; so a question always remains: how to validate the chosen number?

Cluster validation is important because the credibility of the result depends on the number of clusters initially chosen and is important in many medical applications. To address this issue, researchers have proposed different approaches to validate the clusters, or at least give an idea of the number of possible clus-

ters in a given dataset. There are three types of validation studies. An external assessment of validity compares the recovered structure to an *a priori* structure. An internal examination of validity tries to determine whether the structure is intrinsically appropriate for the data. A relative test compares two structures and measures their relative merit [86]. These indices are described in detail in [17, 19, 86, 122].

2.5. Summary

In this chapter, a brief review of existing techniques has been presented, with emphasis on what might be most applicable to cell and nuclei data for the present project. Some basic segmentation approaches are discussed and reviewed briefly in this chapter, in order to understand the proposed segmentation techniques which will be presented next in Chapter 3. The principles of classification techniques such as SVMs are discussed in detail, and other techniques relevant to cell and nuclei classification are also presented, with a brief review of their previous applications.

3. Proposed Segmentation Techniques and Results

3.1. Proposed Segmentation Techniques

The plethora of segmentation techniques, discussed in Chapter 2, is testament to the fact that there is no general approach that performs well across all image types. Moreover, it is difficult for members of the microscopy community to incorporate changes to any one of these methods without a significant investment in effort, and such changes tend to suit only one resolution or segmentation problem. In this chapter, different segmentation techniques are introduced for cells and their nuclei segmentation. Different techniques have been devised, as no single previously developed method worked on all types of data. In the course of this research project, a number of methods were applied, including algorithms that have been previously suggested, including a radial edge search method and a probabilistic boundary optimisation. New algorithms were also developed and tried, specifically aimed at extracting parameters of a model-based fit to endothelial-cell shape. The failure of all these approaches to satisfactory segment a large number of cell boundaries led to an alternative machine learning approach, which is presented at the end of the chapter. In this approach, a user would provide a training set of manually labelled data. The process of automated segmentation would then follow a supervised training, in which an appropriate set of features is first selected, then trained using a suitable machine learning algorithm. These techniques along with the results are presented in this chapter for segmentation for ECs and ECn.

3.2. Radial Method

3.2.1. Edge Map

Initially, the given cells and their respective nuclei images, $f_1(x, y)$ and $f_2(x, y)$, where (x, y) represents spatial coordinates, were passed through respective thresholds (t_1, t_2). For the nuclei image, a global optimum threshold, t_1, was adjusted as: $t_1 = \mu \pm \sigma$, where μ and σ are the mean and standard deviation of intensity values of the particular image.

Connected component labelling [48] was used to enumerate the regions and areas in the binary cell nuclei image, $g_1(x, y)$, with the help of Equation (3.1).

$$\forall i \quad g_i(x, y) = \begin{cases} 1 & \text{if } f_i(x, y) \geq t_i, \quad \forall(x, y) \\ 0 & otherwise \end{cases} \tag{3.1}$$

Here, $i = 1$ or 2. The centroids of the regions were found using moment calculations and a database of cell nuclei was created using the region property values (e.g. areas, centroids, etc). A local threshold, t_2, was set up manually for the cell image according to the intensity values of the candidate cell boundary, in order to get the cell image, $g_2(x, y)$, from Equation (3.1). Here, t_2 varies with the candidate cell. The cell search started from the nuclei centroid position as a reference.

3.2.2. Radial Initialisation

From the centroids of the nuclei and their respective cell images, a radial search was performed using Equation (3.2).

$$h(x_i, y_i) = \forall j : \arg\min\{h(x_i(j), y_i(j))\} \tag{3.2}$$

where $x_i(j) = x_0 + r_i \cos(\theta_j)$, $y_i(j) = y_0 + r_i \sin(\theta_j)$, $i = 0, ..., N - 1$, where N is the number of radial lines, $j = 0, ..., M$, M is the maximum length of the radial lines specified by the user, $\theta_j = 0, ..., (2\pi - \Delta\theta)$ with an increment of $(\Delta\theta = \frac{2\pi}{N})$, (x_0, y_0) is the initial starting point of the radial search and $(x_0$ and $y_0)$ are the centroid location of the nuclei. $h(x, y)$ could be $g_1(x, y)$ or $g_2(x, y)$ depending upon the image on which processing is carried out. In this radial

search $M = 100$ and $N = 512$ radial lines were chosen to get the finer angles of the boundaries. The points which were on the centroid location or those which were very far from the cell were eliminated.

3.2.3. Radial Interpolation

After finding initial boundary points through the radial search, an interpolation was done using the formula given in Equation (3.3). Points found in this way then describe the best potential boundary points in that candidate cell.

$$\forall k : (x_k^{'}, y_k^{'}) = \alpha\beta(x_{k+1}, y_{k+1}) + (1-\alpha)\beta(x_{k+1}, y_k) + \alpha(1-\beta)(x_k, y_{k+1}) + (1-\alpha)(1-\beta)(x_k, y_k)$$

$$(3.3)$$

where α and β are the along x and y axis respectively.

The interpolation function takes initial boundary points, (x_k, y_k), found from the radial search as an input and it also needs to have defined maximum, d_{max}, and minimum, d_{min}, distances between the two nearest boundary points. If the distance between two boundary points, $(k, k+1)$, is less than the minimum distance set, i.e. $(d(k, k+1) < d_{min})$, then a boundary point either at k or $k+1$ is eliminated. If the distance is greater than the maximum distance specified, $(d(k, k+1) > d_{max})$, then new boundary points are added between k and $k+1$.

Figure 3.1 shows the results of radial search on the binarised images (after applying global and local thresholds) of nuclei and boundary images respectively. Figure 3.2 shows the results after applying the radial search method. The algorithms are implemented in *Matlab* version 7.9.0.

The initialisation is done from the centroid of the cell nucleus, which guarantees that it starts within the desired segmentation region. This method works very well even on the concave boundaries; it is also very fast and computationally inexpensive. In this method, manually specified thresholds are needed for cell and nuclei images. This limits its efficacy as the thresholding technique did not recognise cell regions which were below the set threshold. For ECs images, the majority of cell regions were filtered-out, due to weak cell boundaries. For the ECn images, the collaborating biologists did not accept the results of simple thresholding techniques, as they did not describe the nuclear regions

41
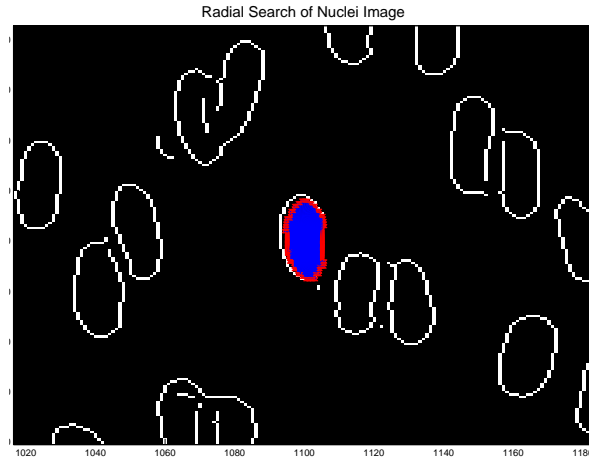
(a) Radial search on ECn.
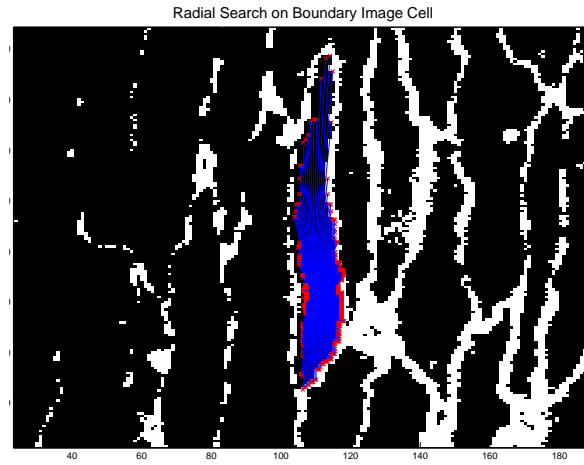
(b) Radial search on ECs.

Figure 3.1.: Radial search results on ECn and ECs edge maps.

42
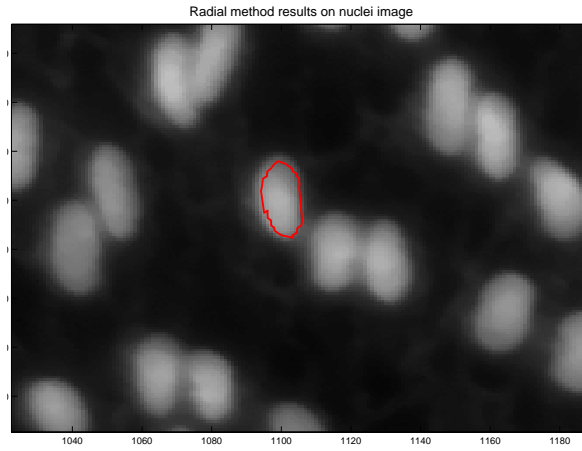
(a) Radial interpolation result on nuclei.
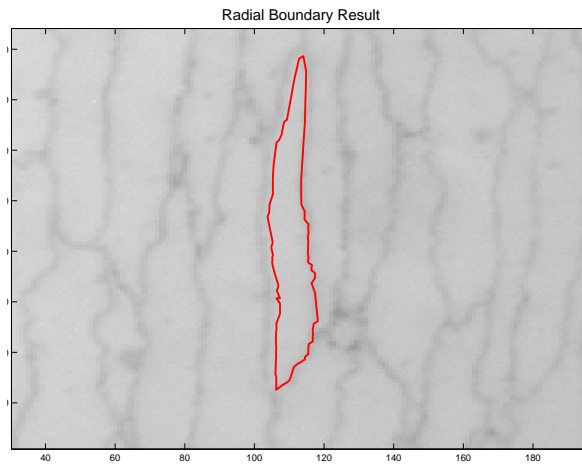
(b) Radial interpolation result on cell.

Figure 3.2.: Radial interpolation results on ECn and ECs images, SNR is sufficiently high in this ECs image.
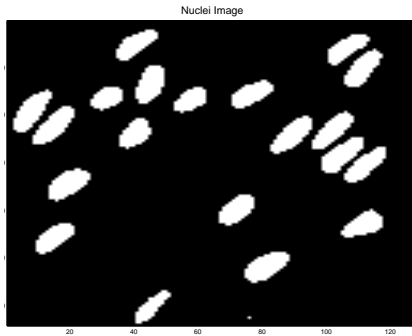
Figure 3.3.: One sample of ECn image after applying **Canny** edge detection and **imfill** methods.

sufficiently well. Therefore, on ECn images, a two-step processing approach was taken which is as follows:

1. Initially, Canny [29] edge detection was applied to the nuclei images. This was performed using the edge-detection method of the *Matlab* image processing toolbox; the default settings for the Canny implementation were used. The output of this function consists of contours of nuclei images (binarised images).

2. Then, on the resulting binary nuclei images (contours of the nuclei) found through step one, a morphological operation (*imfill()*, described in [121], as implemented by the image processing toolbox of *Matlab*) was applied. In this step, a filling of the boundaries of the nuclei edges was achieved, based on 4-connectivity. This operation yields binary images with completely filled nuclei. One example of a binarised nuclei image is given in Figure 3.3.

3.3. Likelihood Maximisation Method

After finding a set of initial boundary points described in the previous section, an algorithm was needed which would optimise these boundary points to get

the most accurate pixel set for EC images. To do so, a likelihood function was defined in with three types of information, intensity values, distance and direction of the candidate pixels, was incorporated. Some of formulations given in this section are based on the prior work [126] of T. Mouroutis of Imperial College London.

When a possible set of boundary pixels $X = \{x_1, x_2, ..., x_n\}$ is known, the likelihood can be estimated that the pixel x_i, $i = 1, 2, ..., n$ belongs to the set X, based on grey scale information. Assume that the conditional probability density functions $p_{grey}(x_i)$ of the intensity values for boundary points from the respective centroid is Gaussian, given by:

$$p_{grey}(x_i|X) = \frac{1}{\sqrt{2\pi} \cdot \sigma_{grey}} \cdot \exp\left(-\frac{(x_{i_{grey}} - \mu_{grey})^2}{2\sigma_{grey}^2}\right) \qquad (3.4)$$

We can evaluate $p_{grey}(x_i \in X) = p_{grey}(x_i)$, where $x_{i_{grey}}$ is the grey scale value of the pixel x_i, μ_{grey} and σ_{grey}^2 are the mean and variance of the p_{grey}. Here, μ_{grey} and σ_{grey}^2 can be calculated from $X'_i = X - x_i$, X'_i is a set of remaining $n - 1$ boundary points, and $\mu_{grey} = E\{X'_{grey}\}$, $\sigma_{grey}^2 = Var\{X'_{grey}\}$.

Following this, the direction vectors of all boundary points from the set of X are calculated. Assume d_i is the direction vector of a current boundary point x_i, then, the six nearest neighbourhoods $N_6(x_i)$ of x_i can be described as $d_i(N_6) = N_6(\overrightarrow{\mu}(x_i))$. The probability of the direction at location (x_i) is given by:

$$p_d(x_i|X) = p_d(x_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma_d} \cdot \exp\left(-\frac{(x_{i_d} - \mu_d)^2}{2\sigma_d^2}\right) \qquad (3.5)$$

The priors of this model are the gradient values. Hence, this function may be written as:

$$p_e(x_i|X) = \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(d(x) - \mu_i)^2}{2\sigma^2}\right) \cdot g_i \qquad (3.6)$$

where N is the number of the gradient image local maxima along the radial search line, μ_i is the distance of the gradient maxima location, σ the variance

of the original Gaussian smoothing filter, and g_i the prior for each Gaussian component of the mixture model calculated as previously described.

After analysing the derivations of the partial likelihood functions for each pixel, the individual likelihood function can be formulated for the ith pixel x_i of the pixel set X. By assuming independence of these partial-likelihood functions, this function can be defined as in Equation (3.7), and the total likelihood function will be calculated as in the Equation (3.8).

$$l(x_i|X) = p_{grey}(x_i|X) \cdot p_d(x_i|X) \cdot p_e(x_i|X) \tag{3.7}$$

$$L = \prod_{i=1}^{n} l(x_i|X) \tag{3.8}$$

3.3.1. Likelihood Maximisation Algorithm

Let $X_0 = \{x_1, x_2, ..., x_n\}$ and $l_i^0 = l(x_i|X_0), i = 1, 2, ..., n$. We define $L(X_0) = \prod_{i=1}^{n} l_i^0$. Let, $j_i = \arg\max l_i^0$ be the index of the pixel with highest value of individual likelihood function. Along the jth radial line a search for the pixel x_m that maximises the value of the total likelihood function $L(X_{modified})$ where $X_{modified}$ is the modified set of pixels created from X_0 by substituting x_j with x_m. Once the pixel x_m has been modified x_j is substituted by x_m in the original X_0. If no pixel has been found along j^{th} search line that increases the value of L compared to $L(X_0)$, pixel x_j remains intact. Then, we repeat the search along the search line corresponding to the pixel with the next highest value of the individual likelihood function. Eventually, all the lines have been searched and the resulting set of pixels is denoted with X_1.

This process, given in Algorithm 3.1, continued iteratively, and after a complete search along the lines, a new, more precise set of pixels is generated. The set created after the kth iteration will be denoted with X_k. The optimization will be terminated when, after an entire iteration, $X_{k+1} = X_k$. The resulting set of pixels X is the set that maximises the likelihood function and is the set of boundary pixels that best describes the potential boundary points of the cell.

In previous studies [126], it was guaranteed convergence of this algorithm. However, when this maximisation method was applied to EC images, it sometimes could not converge, or sometimes it converged on the boundaries of other nearby cells due to distorted boundaries, low boundary signal and poor contrast of the images. Some results are presented in Figure 3.4 on two examples of better quality cells in one image.

Algorithm 3.1 Algorithm for Likelihood Maximisation

$X := [$ initial boundary points $]$ {Initialise boundary points found through radial method described in section 3.2.}
$rank := 0$
$nochange := 1$
while $nochange == 1$ **do**
  $rank := rank + 1$
  $X_{k+1} := L(X_0)(rank)$
  call function for maximising indices
  if $currentboundary == newboundary$ **then**
    $nochange := 1$
  else
    $nochange := 0$
  end if
  $newboundary := currentboundary$
end while

In the next Section 3.4, a demonstration for signal to noise ratio (SNR) and contrast to noise ratio (CNR) of the ECs and ECn images is presented to explain the problem faced during the segmentation methods presented previously in this thesis.

3.4. SNR and CNR of ECs and ECn

To give a demonstration of the SNR conditions, in order to assess the applicability of intensity thresholding techniques, boundary and background points were manually selected from four good quality EC boundary and ECn images, where some of the boundaries were clearly visible and some were missing. The SNR and contrast to noise ratio (CNR) using [22] for the EC boundary and
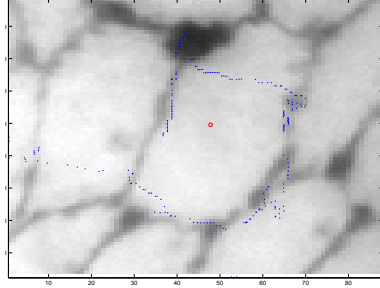
(a) Boundary points results of radial search method described in Section 3.2 on one cell. Note the failure around the cell.
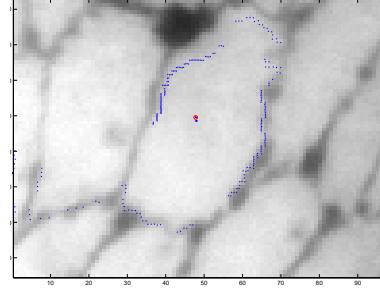
(b) Boundary points results of the same cell in Figure 3.4(a) after likelihood maximisation described in 3.3. Note that some boundary points converge on to a nearby cell.
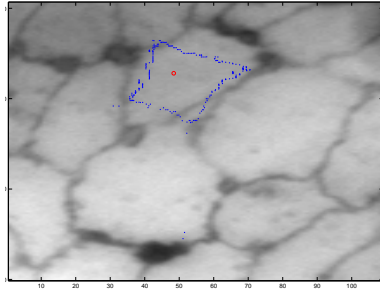
(c) Boundary points results of radial search method described in Section 3.2 on one cell. Note failure to converge to the correct boundary.
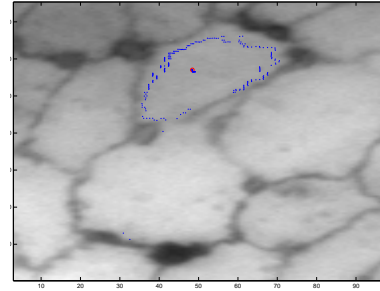
(d) Boundary points results of the same cell in Figure 3.4(c) after likelihood maximisation described in 3.3. Note convergence of points in the center of the cell.

Figure 3.4.: Likelihood maximisation method results for two cells. Cell 1 at top panel and cell 2 at bottom panel. Here, **red o** shows the centroid of the cell nucleus and **blue .** show the boundary points detected.

Table 3.1.: CNR and SNR of EC and ECn images.

		CNR	SNR (dB)
ECs Image	**1**	1.03	0.12
	2	1.21	0.21
	3	0.66	0.31
	4	0.95	0.14
ECn Image	**1**	3.36	12.99
	2	3.24	12.43
	3	2.23	4.63
	4	19.48	9.30

ECn images were calculated by the formulae given below:

$$SNR = 10 \log_{10} \frac{Power_{bd}}{Power_{bg}} \tag{3.9}$$

$$CNR = \frac{\mu_{bg} - \mu_{bd}}{\sqrt{\sigma_{bg}{}^2 + \sigma_{bd}{}^2}} \tag{3.10}$$

where μ_{bg} and μ_{bd} are the averages, σ_{bg} and σ_{bd} are the standard deviations and $Power_{bg}$ and $Power_{bd}$ are the power of the background (bg) and boundary (bd) points in the image respectively. The probability density estimate of boundary and background intensities, as shown in Figure 3.5, is calculated using the class conditional probability distribution functions given as:

$$P(i) = P_{bd}P(i|bd) + P_{bg}P(i|bg) \tag{3.11}$$

where P_{bd} and P_{bg} are the prior probabilities of the pixels and $P(i|bd)$ and $P(i|bg)$ are conditional probabilities of boundary and background respectively. $P(i)$ is the weighted sum of $P(i|bd)$ and $P(i|bg)$ [83], and explains the histograms of all pixel intensities pooled together.

In Table 3.1, it can be seen that SNR and CNR for EC images are very low as compared to ECn images in the same regions. As the SNR decreases, the output pattern becomes noisier [65], and it becomes extremely difficult to detect EC boundaries using conventional intensity-based algorithms developed

for segmentation.
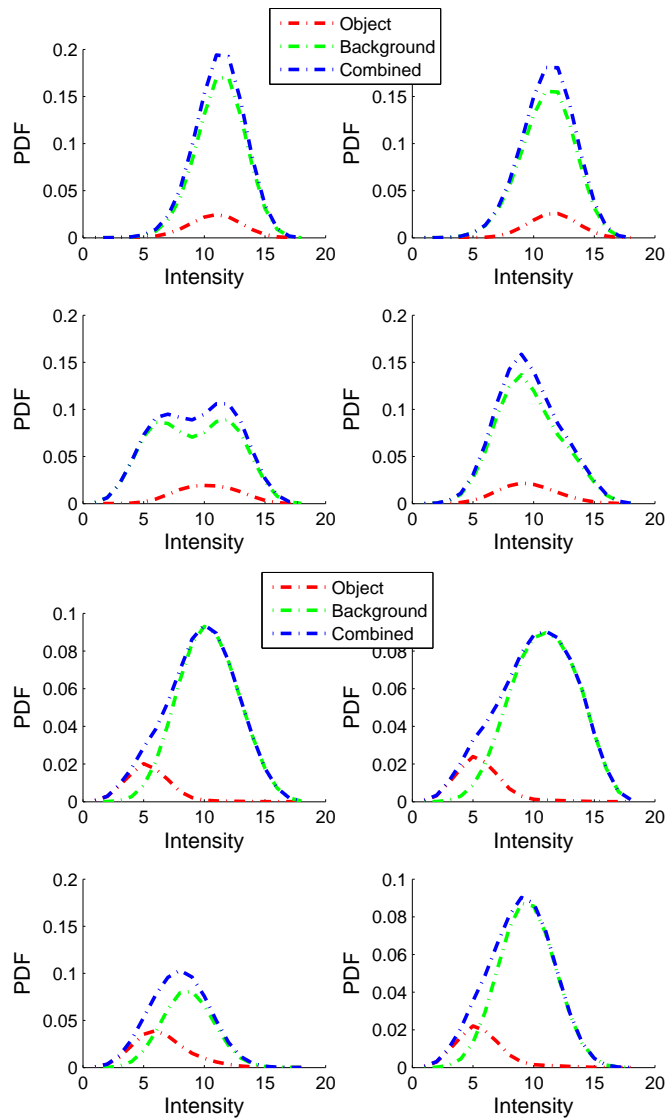
Figure 3.5.: Estimated probability density estimation of intensities of four ECs (top panel) and four ECn (bottom panel) images for object and object and background classes indicating how weak cell boundaries are with respect to the background. Even for ECn images the intensities are not high enough to distinguish nuclei from background easily using intensities only.

Due to the lack of robustness in finding boundary points in EC images using the method described above, we attempted to directly fit an elliptical model to the ECs images in order to overcome SNR problem. Rather than trying to detect the boundaries accurately, an attempt was made to fit a candidate shape model described in the Section 3.5 to the ECs image data. Here it should be noted that, the collaborating biologists specifically wished to find the major and minor axes and angle of orientation of the ECs and their respective ECn to estimate the blood flow and shear stress effects on cells and their nuclei.

3.5. Candidate Cell Shape Model

This section describes a method developed to fit equivalent ellipses to the ECs. In this model, a complex orientation dominant field $\vec{O}(x, y, s)$ of the image $f(x, y)$ can be formulated at a decomposition level of s using Equation (3.12).

$$\vec{O}(x, y, s) = \frac{\sum_{k=1}^{3} |g_{(k\frac{\pi}{4}, s)}| \exp^{(2jk\frac{\pi}{4})}}{\varepsilon + \left(\sum_{k=1}^{3} |g_{(k\frac{\pi}{4}, s)}|^2 \right)^{\frac{1}{2}}} \qquad (3.12)$$

where ε is a constant, $j = \sqrt{-1}$, described in [21], and s is the scale.

This complex field is generated using arrangements of fixed filters, in which addition of vectors of the magnitude of outputs of complex filter units is performed using complex number algebra. The magnitude of the scaled vector responses range from 0 to 1 and can be used as an indication of anisotropy, in which strongly isotropic neighbours will produce values near to 0 and strongly anisotropic neighbours will produce values near to 1 [21]. An illustration of this field is given in Figure 3.6, which has been extracted using the second level decomposition of one of the boundary images.

The cell shape model $\vec{O}_H^{\theta,a,b}(x, y)$ in Cartesian coordinates (x, y) is given by:

$$\vec{O}_H^{\theta,a,b}(x, y) = \left[\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2} \right) * \sqrt{f_x^2 + f_y^2} \cdot \exp\left(i2\theta \right) \right] \qquad (3.13)$$

51
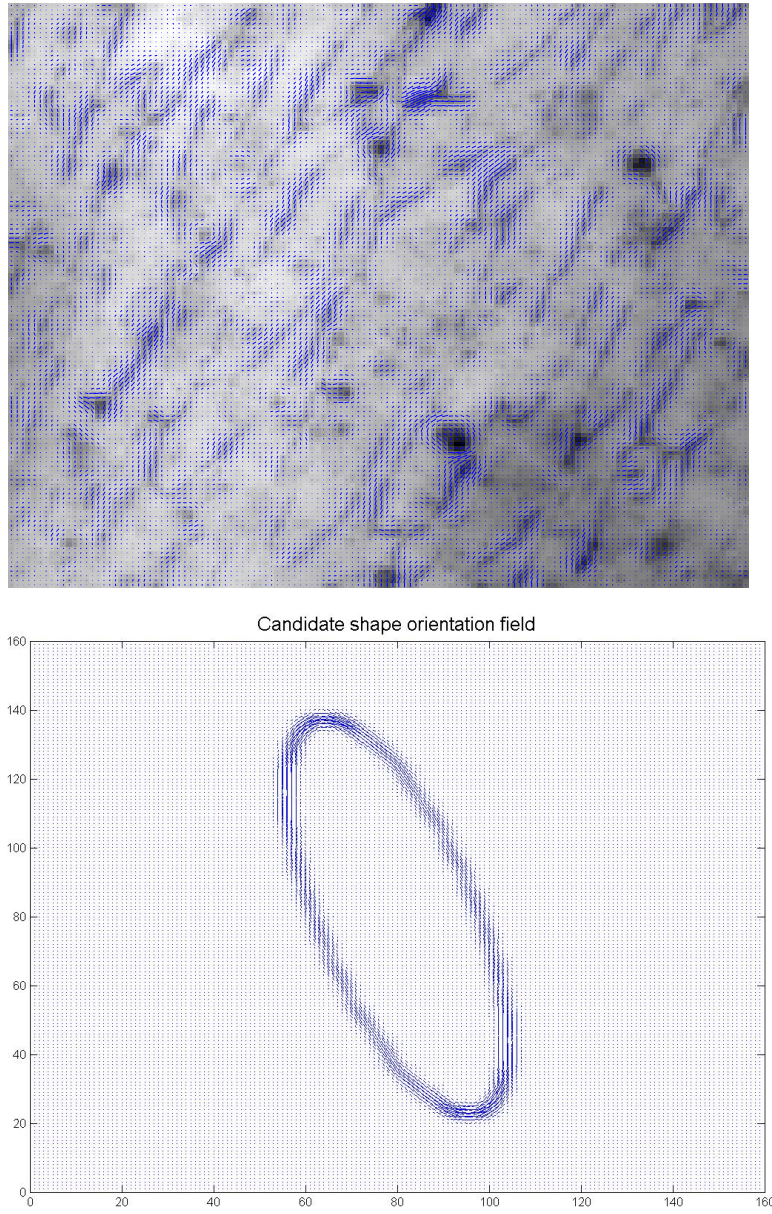
Figure 3.6.: Complex orientated dominance field on original boundary image and a sample template hypothetical orientation field for the candidate cell shape model overlaid in blue lines, generated from the method described in [21], using the second level decomposition.

where $\theta = \tan^{-1}(\frac{f_y}{f_x})$ and $*$ is the convolution operation. This complex hypothetical orientation vector field for the candidate cell shape model is calculated using the equation of the ellipse. The elliptical equation in Cartesian coordinates (x, y) is given by: $f(x, y) = \frac{x - x_0}{a^2} + \frac{y - y_0}{b^2} = 1$, where a and b are the major and minor axes respectively, and x_0 and y_0 are the centroid location of the respective cell. For this, the boundary points of the ellipse with the direction information is calculated by taking the directional derivatives (f_x and f_y) given in Equation (3.14) were convolved with a Gaussian mask to generate a thickened hypothetical orientation field for the candidate shape. The partial derivatives of $f(x, y)$ with respect to x and y are f_x and f_y respectively.

$$
\begin{aligned}
f_x &= b^2(x - x_0) \\
f_y &= a^2(y - y_0)
\end{aligned}
\tag{3.14}
$$

The orientation fields $\vec{O}(x, y)$ of the image $f(x, y)$ given in Equation (3.12) and a hypothetical orientation field $\vec{O}_H(x, y)$ in Equation (3.13) of the candidate cell, are compared by a sum of dot products between vectors of the boundary orientation field of the image $f(x, y)$ space and vectors of the hypothetical orientation field for the given shape around that candidate cell.

The similarity of these fits is optimised over parameters (orientation θ, major and minor axes a and b) given in Equations (3.15 and 3.16). The whole process is summarized in the flow chart given in Figure 3.7.

$$
\frac{\int_{Boundary} \vec{O}(x, y) \bullet \vec{O}_H(x, y | \theta_0, a_0, b_0; x_0, y_0) \; dx \; dy}{|\int_{Boundary} \vec{O}_H(x, y)|}
\tag{3.15}
$$

To directly fit the parameters sought (a, b, and θ) centred on the nuclei centroid at (x_0, y_0) to begin with the fit of the ellipse. We then perform the following:

$$
\{\theta_0, a_0, b_0\} = \arg\max_{\theta, a, b} \frac{\int_{Boundary} \vec{O}(x, y) \bullet \vec{O}_H(x, y | \theta, a, b) \; dx \; dy}{|\int_{Boundary} \vec{O}_H(x, y)|}
\tag{3.16}
$$
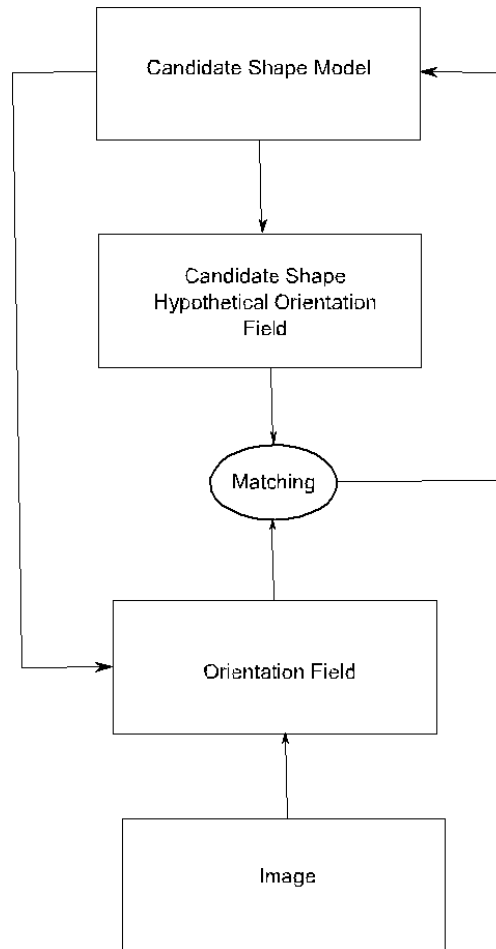
Figure 3.7.: Flow chart for the candidate cell shape model.
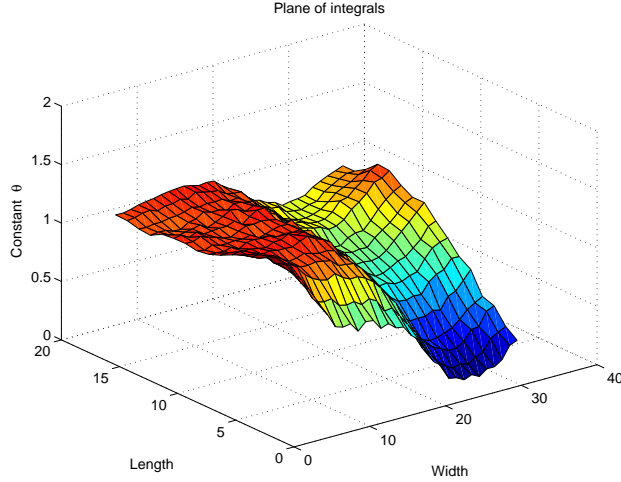
Figure 3.8.: A sample plane of integrals over major, minor axes and orientations.

A template for the extracted hypothetical orientation field is given in Figure 3.6. This vector field is generated using second level decomposition, with a Gaussian mask of 7 by 7 and depends on the parameters θ, a and b. The template is the same size as the given image patch. Then, using the orientation field of the image generated, as described in [21], and the hypothetical field of the candidate cell (i.e. an ellipse), as shown in Figure 3.6, is optimised over angle, major axis and minor axis of the shape of the candidate cell. In this study, the major and minor axes (a and b) values were chosen with an initial estimate and were measured in pixel units. A range of values between 50 to 110 and between 10 and 35 were chosen for a and b parameters respectively; and θ varies between 0 to π. One example of a plane of integrals is presented in Figure 3.8. Here, the best fit to candidate shape or cell was chosen when a combination of parameters θ, a and b is maximised.

This section has described the model developed for fitting the equivalent ellipses on ECs boundaries. This model worked well on a few cells where the cell boundaries were clear, but wherever the cell boundaries were distorted due to noise or image acquisition problems or natural distortions, fitting of the ellipses

55
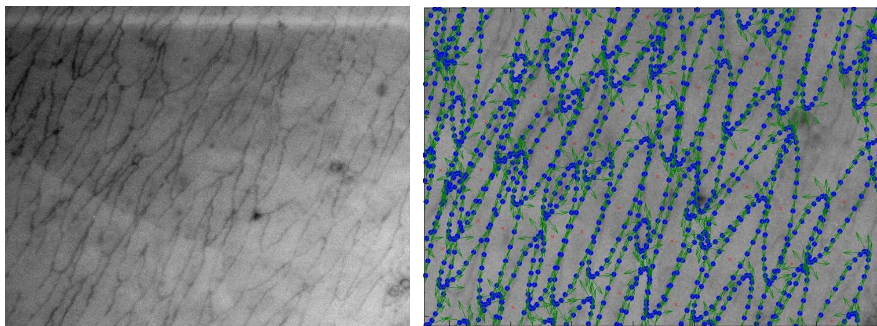
Figure 3.9.: Fitting of ellipses on original image with centroids marked in red ✕.

failed. One example of a good performance of an ellipse-fit with its original image is shown in Figure 3.9. Evaluation of different methods presented till now is presented in the next Section.

3.6. Evaluation of Ellipse-fits on Different Proposed Segmentation Techniques

To evaluate the method, different users were asked to fit ellipses manually to the cells where they can be seen clearly, given the centroid locations of the cells from the nuclei images. Eight naïve human users fitted ellipses on 3 different images, and cells were chosen which generated similar fits in all the 8 users. These cells were used as *ground-truth* for the evaluation of the methods described so far in this chapter, the average and average ±1 SD for each of the *ground-truth* cells was found and overlaid in blue thick and dashed ellipses in Figure 3.10. It can be seen even in the ground truth results, that there was great variability in the fitting of the ellipses by the different users.

Next, using the results found through the radial method described in Section 3.2, an *ellipse-fit*() function, available online in *Matlab* file exchange [9], was applied to find the θ, a and b of the cells, and the candidate shape cell model ellipse results were overlaid on the corresponding cells. These elliptical parameters found from different cells are given in Table 3.2. It can be observed that

none of the methods, presented until now, worked well on all cells.
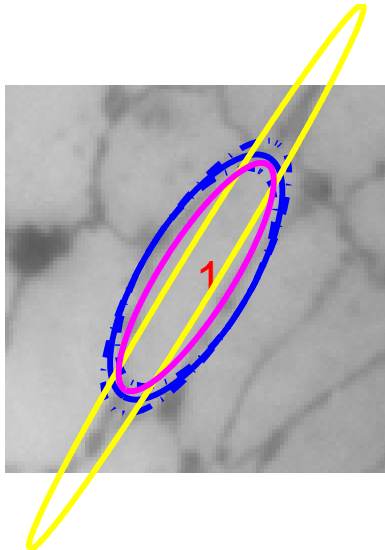
Figure 3.10.: Ellipses overlaid on the original cell. Ellipses in blue, -.-.- blue, yellow, and pink represent *ground-truth*, mean ± 1 SD *ground-truth*, candidate shape model and radial method segmentation techniques respectively.

It was therefore decided to apply machine learning algorithms to detect the cell boundaries, and to do so, SVMs were invoked to perform cell boundary detection. The motivation to use SVMs [171] for segmentation is explained in the literature review, because recent work has shown this approach to outperform competing methods in many applications [51].

3.7. Segmentation Algorithm using SVMs

This section presents an application of the SVM learning as a classifier with a sufficiently rich feature vector to detect thin boundaries of the ECs. The flow chart of our algorithm is given in Figure 3.11. Features are extracted from the cell boundary image and mapped into feature space. To train the classifier, features are extracted from a training image for which there exists manually labelled data. Then, using the input feature space generated from the training

Table 3.2.: Various ellipse parameters for *ground-truth*, 1SD *ground-truth*, candidate shape model and radial method segmentation techniques respectively.

Major Axis (Pixels)	Cell #1	Cell #2	Cell #3	Cell #4	Cell #5	Cell #6	Cell #7	Cell #8	Cell #9
Mean	35.07	30.75	39.05	38.84	55.69	54.04	50.24	53.95	65.85
1×SD	4.13	4.06	6.85	8.98	13.70	12.47	13.58	23.18	28.80
Candidate Shape Model	45.00	95.00	120.00	115.00	90.00	50.00	50.00	70.00	75.00
Radial Method	24.02	59.25	76.79	34.16	13.63	26.78	114.13	12.81	40.87

Minor Axis (Pixels)	Cell #1	Cell #2	Cell #3	Cell #4	Cell #5	Cell #6	Cell #7	Cell #8	Cell #9
Mean	13.51	13.77	14.41	12.50	10.66	13.30	12.63	10.42	9.83
1×SD	2.28	2.45	1.58	2.06	1.72	2.04	2.42	2.58	1.19
Candidate Shape Model	29.00	5.00	12.00	5.00	5.00	6.00	5.00	26.00	5.00
Radial Method	7.11	50.27	54.19	14.39	7.02	8.45	6.81	5.03	7.53

Orientation (Degrees)	Cell #1	Cell #2	Cell #3	Cell #4	Cell #5	Cell #6	Cell #7	Cell #8	Cell #9
Mean	-1.03	-0.72	-1.01	-1.32	-1.08	-0.93	-1.29	-0.91	0.76
1×SD	0.09	0.10	0.09	0.05	1.05	1.01	0.04	1.00	1.43
Candidate Shape Model	0.23	-0.20	-0.87	-0.22	-0.49	0.02	-0.45	-0.36	0.17
Radial Method	-1.06	0.98	0.04	-1.29	-1.44	-1.38	-1.55	-1.40	-1.21

image and an SVM kernel, a decision model is generated; this model is then applied to segment other boundary images. The SVM classifies image pixels into two different classes: boundary and non-boundary.
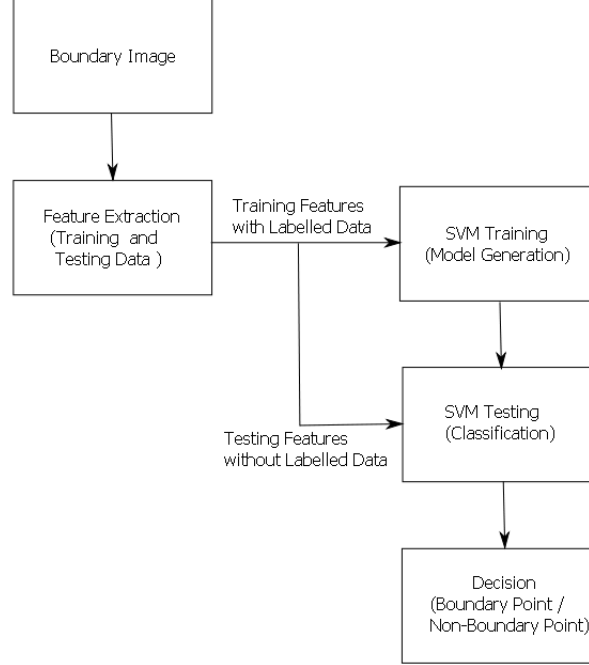
Figure 3.11.: Flow chart of SVM-based segmentation algorithm for endothelial cell images.

3.7.1. Feature Selection of ECs Images

It was found that cell boundaries were either missing from some parts of the cell or were thin and of low contrast, making it visually challenging to distinguish between background and cell boundary. To address this requires the use of features that take into account the structure of neighbourhood pixels. Although low-level edge-linking is usually applied for these types of tasks, it is also possible to use a sufficiently rich feature vector that incorporates neighbourhood information appropriate for describing the weak, thin nature of cell boundaries.

SVMs, as classifiers can perform well in high-dimensional datasets [51]. The-

refore, the raw neighbourhood pixel values over a 3×3 pixel neighbourhood of a central pixel were used as feature components, giving a 9 dimensional feature vector for each pixel. This 3×3 pixel grid was chosen as being typical of published medical image detection applications (e.g. as in [51]) using SVMs. The best possible SVM hyper-parameters were then found using a set of hand-labelled *ground-truth* data. By adding in and removing features, searching over SVM hyper-parameters, such as v (tolerance), and γ radial basis function (RBF) width, an optimal set of features was exhaustively found in each case. The feature extraction process is next described for the set of features that is found to lead to a better classification performance.

3.7.2. Feature Descriptor

The general motivation and an intuitive explanation for each selected feature are explained in this section. Different derived measurements are estimated around each pixel of an image and assigned to an **d** dimensional feature vector. In this study, 33 features are extracted from the neighbourhood of the central pixel. Although some of the features appear to be constructed from a 3×3 pixel neighbourhood, those features that are based on filtering, such as gradient fields and orientation dominance fields, in practice utilise information from much wider neighbourhoods. The selected features include:

Each pixel in the image is mapped to a 33 dimensional feature vector by extracting information from a 3×3 neighbourhood of the central pixel.

- Nine intensity elements extracted from 3×3 nearest neighbours (**F1-F9**).

- Eighteen features, (**F10-F27**), extracted from 3×3 nearest neighbours of an orientation dominance field [21]. The magnitude of the vectors in this field ranges from 0 to 1. The magnitude of the values can be used as an indication of anisotropy, in which strongly isotropic neighbours will produce values near to 0 and strongly anisotropic neighbours will produce values near to 1; an example of this field over a 3×3 patch is shown in Figure 3.12.

- Six statistical features generated from the 3×3 intensity neighbourhood of a candidate pixel, (**F28**-**F33**), given in Table 3.3.

Table 3.3.: Features generated from the 3×3 intensity neighbourhood of a candidate pixel.

Feature Number	Feature Name	Equation
F1-9	**Intensity neighbourhood**	$\left(I_{(i)\sqrt{n}\times\sqrt{n}}(x,y) \right)_{i=1:n}$
F10-27	**Orientation neighbourhood**	$\left(\vec{O}_{(i)\sqrt{n}\times\sqrt{n}}(x,y,s) \right)_{i=1:n}$ $s = \text{scale.}$
F28	**Median**	$Med(x,y) = med\left(I_{(i)\sqrt{n}\times\sqrt{n}}(x,y) \right)$
F29	**Range**	$Range(x,y) = [Max(x,y) - Min(x,y)]$ where $Max(x,y) = max\left(I_{(i)\sqrt{n}\times\sqrt{n}}(x,y) \right)$ $Min(x,y) = min\left(I_{(i)\sqrt{n}\times\sqrt{n}}(x,y) \right)$
F30	**Energy**	$E(x,y) = \sum_{i=1}^{n} I^2_{(i)\sqrt{n}\times\sqrt{n}}(x,y)$
F31-33	**Second, Third and Fourth order moments**	$M_r = \frac{1}{n}\sum_{i=1}^{n}\left(I_{(i)\sqrt{n}\times\sqrt{n}}(x,y) - \mu(x,y) \right)^r$ $\mu(x,y) = \frac{1}{n}\sum_{i=1}^{n} I_{(i)\sqrt{n}\times\sqrt{n}}(x,y),$ $I(x,y)$ is the image, $r = 2,3,4$: second, third and fourth order moments, n is the number of nearest neighbors (in our case this is 9), and i is the candidate pixel.

The 33-dimensional feature vector given in Equation (3.17) is then described for each pixel in the image in both the training and testing phases of the SVM.

$$\mathbf{FD} \;=\; [I_{(3\times3)}, \vec{O}_{(3\times3)}, Med, Range, E, M_{2,3,4}] \qquad (3.17)$$
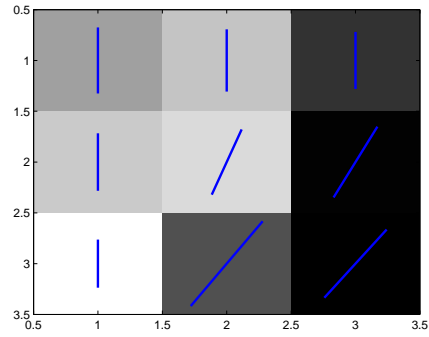
Figure 3.12.: Orientation-dominance field overlaid in blue within a 3×3 neighbourhood of a candidate pixel.
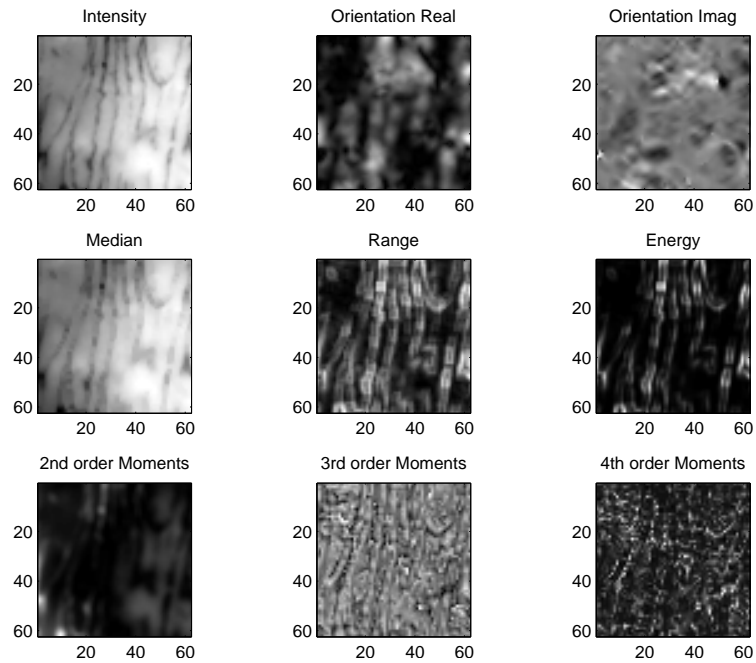
Figure 3.13.: Examples of different features in the training sample.
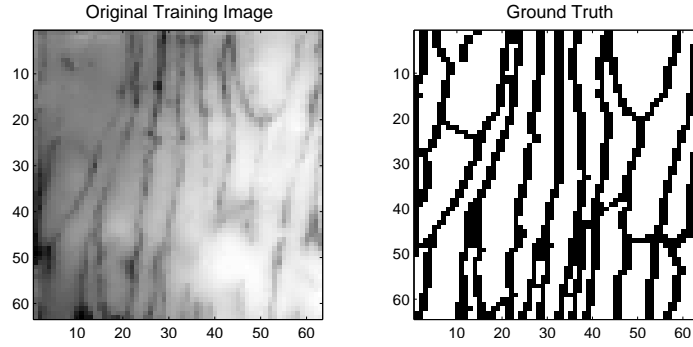
Figure 3.14.: Original image with its ground truth.

Before classifying, the feature vectors were z-scale transformed, using the formulation, given in Equation (2.2), to get a normalized feature space with a zero mean and a unit variance for each component. Some sample training features are shown in Figure 3.13.

3.7.3. Detection of ECs using an SVM-based Method

The normalised 33-dimensional feature vectors are used as an input space to a supervised SVM classifier that classifies image pixels into boundary and non-boundary pixels. A binary image is generated in classification, with each black pixel corresponding to a boundary and each white pixel corresponding to a non-boundary pixel [63, 98].

The SVM is first trained using the training input feature space with its *ground-truth* as *a priori*, and a kernel to generate a classification model. This model is then applied to segment (classify) other "unseen" images of the same general class. This process assigns image points to one of two different classes: object and non-object. The training patch and its *ground-truth* are shown in Figure 3.14.

An RBF SVM kernel, given in Table 2.1, was applied to generate a model, using LIBSVM [32] via its *Matlab* interface. The SVM model performance is dependant on the selection of the model parameters; therefore, during training, the parameter space was swept to find those parameter choices that yielded the best accuracy in the *training* data. This is computationally expensive, as the

search tries values of each parameter across a wide range. As, the RBF kernel mainly uses two parameters (C and γ), a range of -4 to +4 with interval of 0.1 for C, and a γ range from 0.1 to 2 with an interval of 0.001 were selected. To optimise these parameters, 5-fold cross-validation was done on the training image patch, producing 2144 support vectors. The feature extraction process for the training phase took approximately 130 seconds. The model generation process then took 13.38 seconds on a desktop PC with Pentium(R)4 2.8 GHz processor running Windows XP.

Once the model with the smallest error is generated, segmentation is then achieved in a straightforward manner. The normalised 33-dimensional feature vectors of the image to be segmented are computed and supplied to the trained SVM classifier. This results in a labelled image where each pixel has a value depending on its class (boundary and background). The segmentation algorithm was applied to images of 56 regions around 7 branches in aortas from 3 different animals. The time in the classification phase varies with the size of the images to be segmented, but is approximately 12-14 seconds if the features have already been extracted.

The algorithm detected boundaries to a usable degree of accuracy in 43 out of 56 regions. In the remaining 13 regions, the image data was blurred or noisy; one such image is shown in Figure 3.15, and it may be noted that even discerning the boundaries by eye is difficult. The binary images in Figure 3.16 show more typical classification results of pixels into boundaries and background by the SVM algorithm.

3.7.4. Evaluation of SVM-based Segmentation Method

The performance of the SVM-based algorithm in segmenting cell boundaries was evaluated on four images containing manually traced boundaries. In total, the performance of the algorithm was assessed using hand-lablled ground truth on 3.9 million pixels, and the best-performing kernel ranged from 81% to 94% accuracy.

In another experiment, four image patches of size 128×128 pixels with their 'ground truth', an SVM RBF kernel and two conventional segmentation algo-
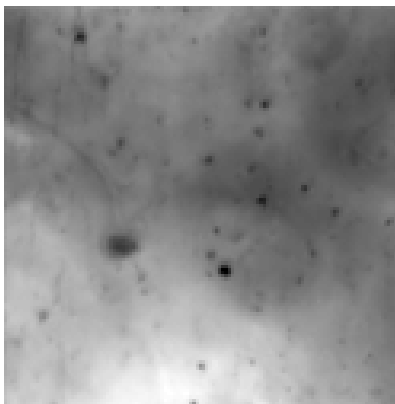
Figure 3.15.: An example of a blurred image (256 × 256 pixels) on which the SVM technique did not work well. Note that the cell boundaries are not discernable.

rithms, Canny [29] and Kittler [96], were compared. The Kittler method was ranked as the best in one thresholding survey paper [150]. Different thresholds were set for the decision in order to obtain a number of points along a Receiver Operating Characteristic curve, which describes the performance of a classification method and feature space together in addressing a discrimination task as one alters the balance between either type I errors or type II errors. For the Canny, 961 possible detection thresholds were applied, and for Kittler, 63 possible thresholds were applied to each image. The SVM model accuracy is largely based on the selection of the model parameters, so to find points along an ROC, a search algorithm attempts training and classification where values of the most significant kernel parameters are sampled across a wide range. However, since each image to be assessed might have different statistical characteristics, generating an ROC might require slightly different parameters settings. For a given image, the model is further evaluated with 5-fold cross-validation to find the best possible combination of parameters. Table 3.4 and Figure 3.17 presents the classical segmentation technique results in comparison to SVM RBF kernel, and the Receiver Operator Characteristics (ROC) plots are given in Figure 3.18. Table 3.4 also includes a comparison with the Otsu method [132],
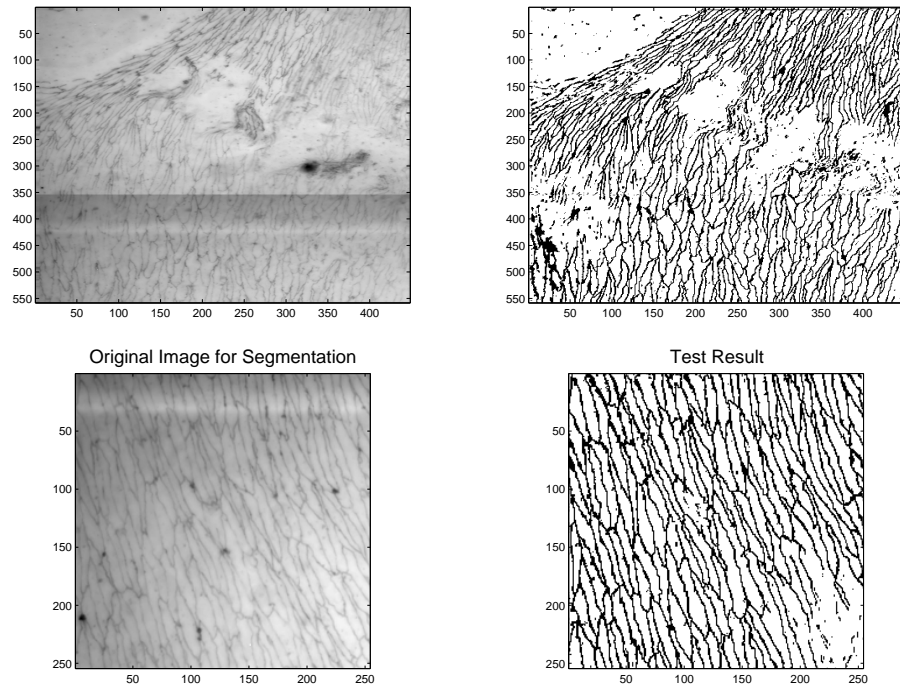
Figure 3.16.: Original image and segmented result using RBF kernel. No other processing has been applied to the SVM classification output and results are shown at half resolution.

Table 3.4.: Comparison between SVM and different conventional methods. Here the number of training pixels was 3844 with 2 classes, and the number of test pixels was 16384.

Method	Accuracy
SVM with RBF Kernel	94.09
Otsu [132]	82.22
Kittler [96]	82.21
Canny [29]	71.83

which employs an automatic thresholding and is therefore not easily amenable to ROC analysis.
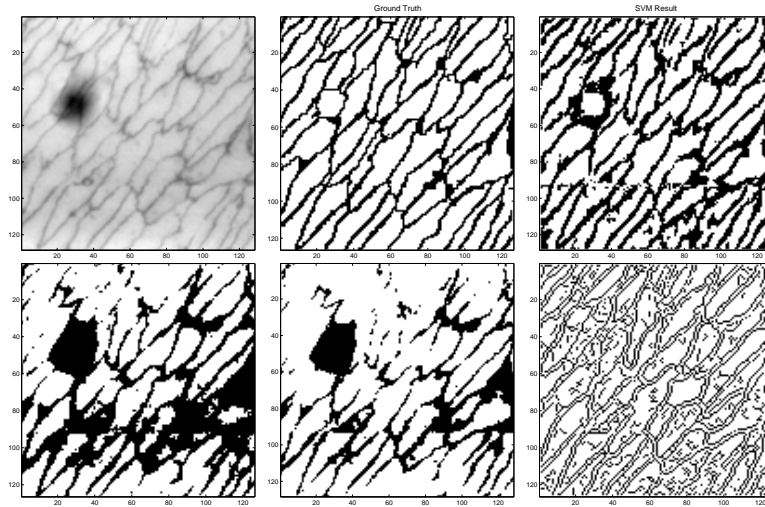
Figure 3.17.: Comparison of results from different image segmentation techniques. Top panel, original image with its ground truth and SVM segmentation results. Bottom panel, Kittler method, Otsu method and Canny edge detection results.

Accuracy Measures

The accuracy of segmentation was assessed by noting the number of correctly classified pixels (either boundary or background) and is defined by (correctly
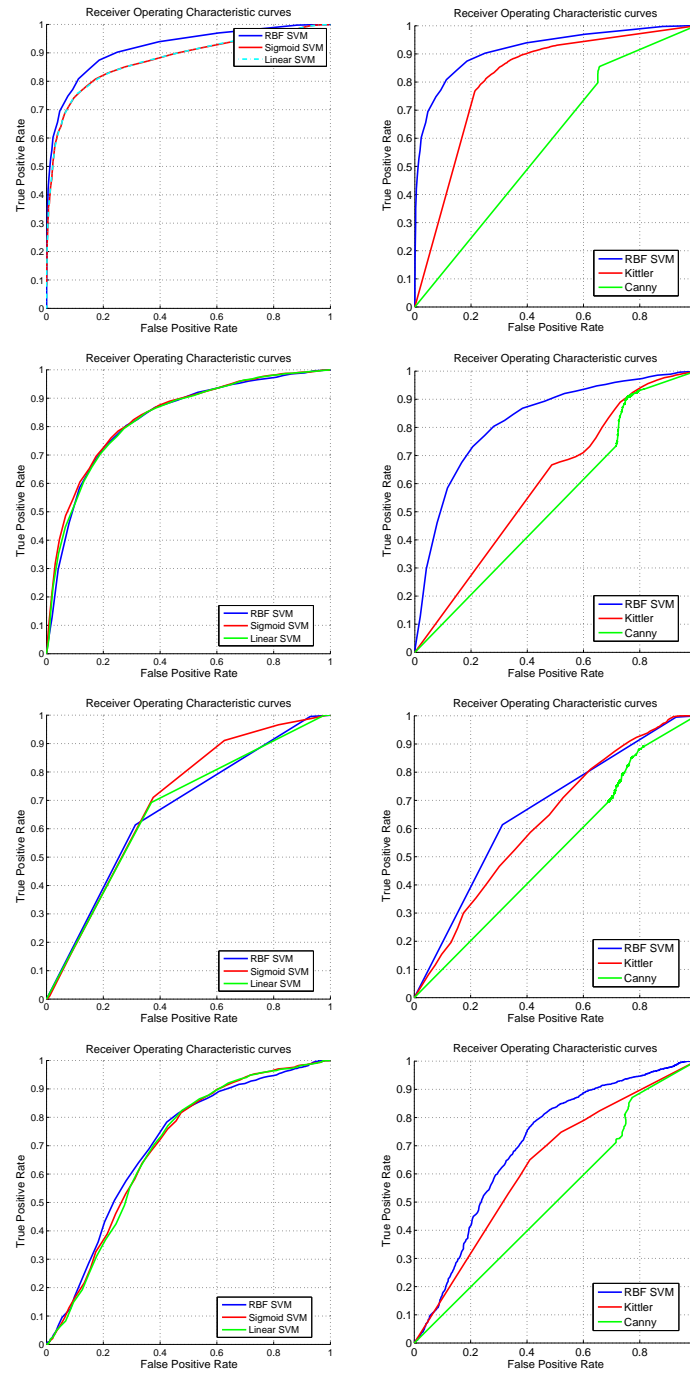
Figure 3.18.: ROC plots for different standard SVM kernels available and Canny and Kittler methods on four images. The performance of the SVM is better, even when applied to the extremely difficult last two images.
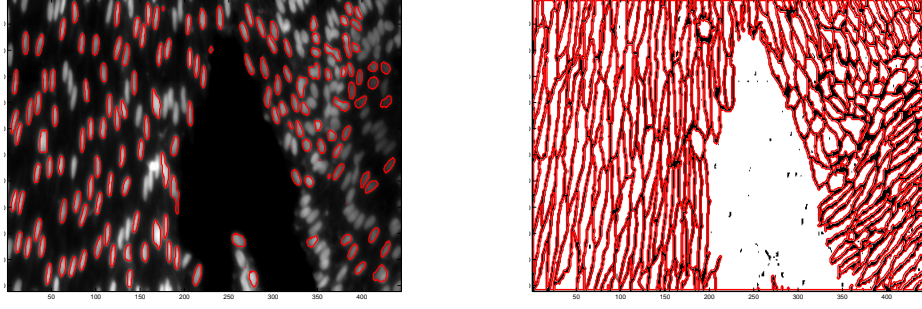
Figure 3.19.: One sample of cell and nuclei image pair with traced boundaries overlaid in red using boundary tracing method described in Section 3.8.

classified pixels):

$$Accuracy\% = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3.18}$$

where TP,FP, TN, and FN are true positive, false positive, true negative and false negative respectively.

3.8. Cell Boundary Tracing

After finding the binary images of the ECs using the SVM algorithm and those for the ECn using the Canny method described in Section 3.2.3, the function *bwtraceboundary*() in *Matlab* image processing toolbox was applied to trace the boundaries of the segmentation regions. This function results in a structure array of size n, where n is the number of objects in the binary image. Each element in the structure contains an X-by-2 matrix, and each column represents x and y- coordinates of the boundary pixels. This function traces the object boundaries all the way round and the number of rows depends on the size of the object. One sample pair of ECs and ECn images with traced boundaries is presented in Figure 3.19.

3.9. Discussion

In this chapter, several methods for detecting the ECs and ECn are presented. These techniques reduce the need for human intervention. Previously this process had been done manually [38, 24], since noise and unclear or faint boundaries lead to the failure of traditional automated segmentation techniques. Manual analysis is time consuming, subjective and unscaleable. Previous reports have described computational work on corneal endothelial cell boundary detection, such as [62, 149, 105, 175], but the cell boundaries are not as challenging to detect as the EC boundaries in our case. In the initial methods proposed, limitations were revealed, such as the problem of correctly finding an initial edge map in Section 3.2 and failing to fit the ellipse properly in the method described in Section 3.5. To address these, a method was proposed utilising a supervised SVM. This was successfully applied to image data with a poor signal-to-noise ratio. This algorithm takes into account small spatial centre-surround differences in intensity, and orientation dominance patches and other properties of each candidate pixel in the image, in both training and classification phases, successfully segmenting 43 out of 56 available regions to an acceptable level (94 %) of accuracy. In this way, an adequate number of cells were collected for further automatic morphological analysis.

This method does have some limitations as it does not perform well on image patches where the brightness variation is large and cell boundaries are not visible by eye. It also requires a manual labelling of one example as input in the training process. This necessitates selecting a patch with a representative variation in intensity to yield a satisfactory generalised model. Another aspect of our algorithm which could be improved is the selection of kernels [39]. In this study, work has been done only on kernels already available in a standard SVM package, but performance might be improved by developing a new kernel for such data, which was beyond the scope of this research. Nevertheless, the method will enable to explore the relationships between blood flow and cell behaviour, between cell nuclear shape and cell boundary shape, and between aspect ratio and cell orientation in different regions around arterial branches, by allowing large numbers of cells to be analysed, which will be presented in the next Chapter 4.

4. Comparative Analysis of *In-vivo* Endothelial Cell Data

4.1. Introduction

In this chapter, we present the statistical analysis of changes in patterns for stress effected morphological parameters of ECs and ECn with age in rabbits. These statistics were obtained using the binarised images found from the segmentation techniques previously presented in Chapter 3.

4.2. Correspondence between ECs and ECn

This section describes an algorithm which we have developed for supervised region-wise matching for two binarised images of cells and their nuclei acquired with two different imaging modalities based on their morphological measurements. This algorithm first created two cell databases, one each for cells and their nuclei images.

After obtaining the binarised images for both ECs and ECn images, it is necessary to establish the correspondence between ECn and ECs for further comparative analysis. In this study, with the help of prior information about cells and nuclei regions, region-wise matching is done on each database for the ECs and ECn, resulting in a complete set of cells and their corresponding nuclei regions. The results presented here for 56 sets of images (ECs and ECn) demonstrate that the algorithm found exact matching of cell regions automatically, which provides a starting point for further automatic analysis of collections of cell regions and their respective nuclei. The algorithm is presented in the next section.

Since segmentation methods are never 100% accurate, establishing correspondence in labelled regions between ECn and EC images is not necessarily straight forward. For example, for any one cell, either the boundary may be missed or merged with a neighbouring cell, and so entire cells may lack corresponding structure in the image pair. This may lead to incorrectly paired nuclei and boundaries, as shown in one example in Figure 4.1, where nucleus does not exist, even then a cell is detected, making a study of the relationship in shear-stress dependent parameters difficult.

From the binary images, all objects are identified using the connected component labelling method, available in *Matlab*. This process is known as labelling, in this each individual object is named with a unique identifier.

4.2.1. Association between ECs and ECn

As it is already known that the cell nucleus must be within the candidate cell. Using this as *a priori* knowledge, the algorithm searched for the best matches between cells and their nuclei; a flow chart is given in Figure 4.2.

With the help of the nuclei centroids, the regions (labelled components) in the cell images were located, and the remaining regions were discarded. Only these found regions were then added into the new ECs and ECn databases. On these regions in the new databases, several conditions were set to ensure their validity. For example, if a region was within a certain range of parameters (such as: area, length, and width), then that region was added to the updated database of cells, otherwise it was again discarded. In the same way, a search is applied to the nuclei regions to find the corresponding boundary region for the nuclei. The algorithm for the association between ECs and ECn is given in Algorithm 4.1. Here, the minimum and maximum values of area, length, width values were chosen from studies covered in the literature review. These values specify a search range used to establish correspondence. Some values are given in Tables 4.1 and 4.2. One pair of ECs and ECn images is presented in Figure 4.1 after applying this algorithm. As we can see in Figure 4.1, the nuclei and their respective cells are located and matched, which produces a set of database for nuclei and their respective cells (1496 cells, in total).
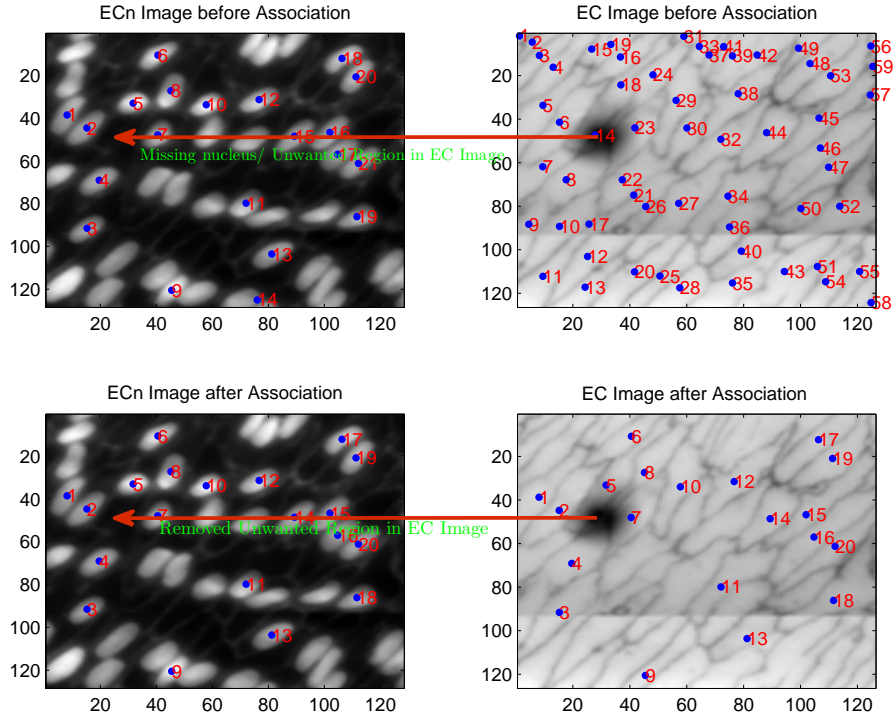
Figure 4.1.: On top, the result of an ECn and EC image pair (labelled with their
selected regions) after the connected components method is applied
initially. Numbers represent the centroid locations of the regions.
Note the missing nucleus or unwanted cell region. Bottom, result
after applying the matching algorithm described in Algorithm 4.1
on the ECn and EC image pair. Numbers represent the centroid
locations of the regions. Note that the unpaired cell regions are
removed from images.

Investigator	EC Length μm	EC Width μm	EC Area μm^2
Cornhill [38]	63	19	N/A
Nerem [73]	66.2	14.0	N/A
Lavesque [111]	42.7	20.5	N/A
Andy [24]	39.7	14.4	488.9
Average	52.9	16.9	489

Table 4.1.: Average ECs values reported in literature. Adapted from [24].
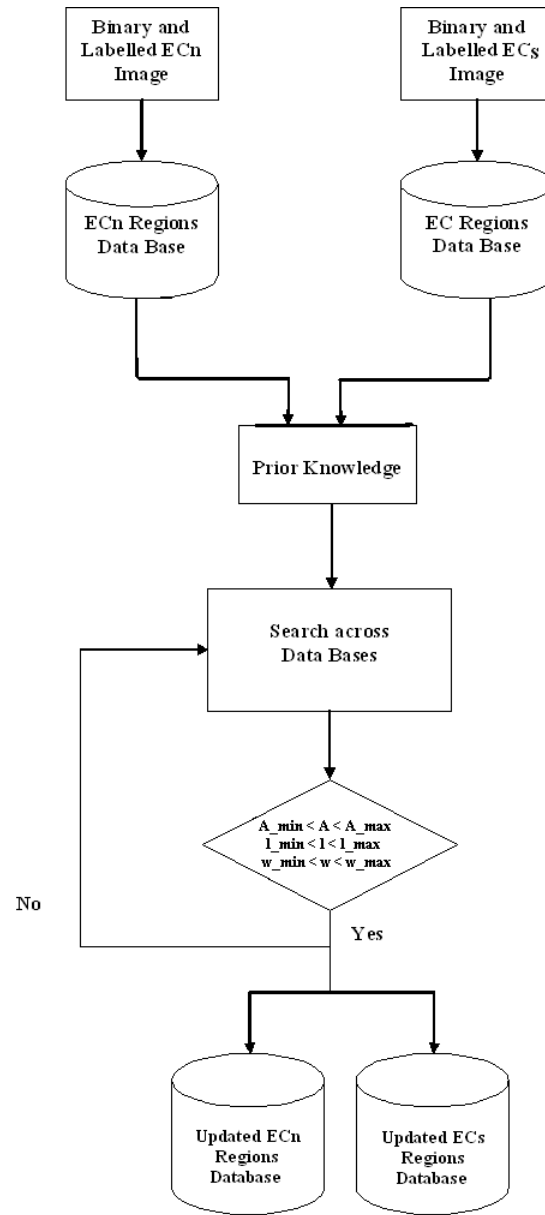
Figure 4.2.: Flow chart for finding the correspondence between cells and their nuclei in separate image databases.

Algorithm 4.1 Association between ECs and ECn

Get labelled nuclei and cell images
find centroids of nuclei and cell regions
$n :=$ no: of nuclei regions
for $i := 1$ to n **do**
   find labelled cell region at locations of ith region of nuclei
   Add found entry into a new ECs database
end for
For both ECs and ECn values given in Tables 4.1 and 4.2
select MaxA, MinA, MaxL, MinL, MaxW, MinW
$m :=$ no: of cell regions , $m \leq n$;
for $j := 1$ to m **do**
   *function***CellRegionConditionalValidityCheck**(CellRegion(j))
   if $MinA < CellArea < MaxA$ **then**
     if $MinL <$ CellMajorAxis¡$MaxL$ **then**
       if $MinW <$ CellMinorAxis¡$MaxW$ **then**
         Create a new database with conditional ECs and ECn regions
       end if
     end if
   end if
   return(ValidCellRegion)
   Add validated entry into new ECs and ECn databases
end for

Investigator	ECn Length μm	ECn Width μm	ECn Area μm^2
Al-musawi **Immature**	645 ± 56	573 ± 29	N/A
[16] **Mature**	703 ± 51	543 ± 106	N/A
Andy **Immature**	16.53 ± 0.02	7.10 ± 0.02	489.59 ± 1.6
[24] **Mature**	18.70 ± 0.03	6.01 ± 0.01	483.90 ± 1.21
Average			487

Table 4.2.: Average ECn values reported in literature.

4.2.2. Morphological Measures

From each object, different shape measures can be calculated for further analysis, such as length, width, and angle. Other measures include:

The general equation for moments calculation is defined as:

$$M_{pq}(x,y) = \sum_{i=1}^{n} (x_i - \mu_x)^p (y_i - \mu_y)^q I_{(i)}(x,y) \tag{4.1}$$

Here (x,y) are the pixel coordinates of I (image), $\mu_x = \frac{1}{n} \sum_{i=1}^{n} I_i(x,y) \times x_i$,

and $\mu_y = \frac{1}{n} \sum_{i=1}^{n} I_i(x,y) \times y_i$.

Centroids

The centroids or zeroth order moments or central moments can be calculated from moments calculations as: $\mu_{pq} = \frac{M_{pq}}{M_{00}}, \quad p+q = 1$.

Length(L)

The longest distance from the centroid to the edge of the object is known as the long-axis or a measurement for length of an object.

Width (W)

The shortest distance between boundary points and the centroid of the object is termed the short-axis or width of an object.

L:W Ratio

This is the ratio between long and short-axes. It is a measure of elongation.

$$L : W\,Ratio = \tfrac{L}{W}$$

Angle(Θ)

The angle is the measure of the degree at which the tangent of the object and the long-axis(L) intersect.

$$\Theta_i = \tfrac{1}{2} \arctan \left(\tfrac{2M_{11}}{M_{20}-M_{02}} \right).$$

where M_{pq} (moments) general equation is defined in Equation (4.1).
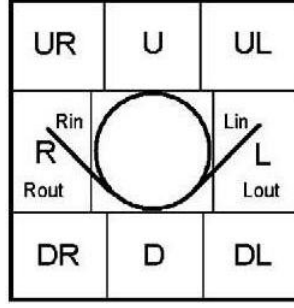
Figure 4.3.: Map for eight vascular regions, here Upstream Right (UR), Upstream (U), Upstream Left (UL), Right (R), Left (L), Downstream Right (DR), Downstream (D), Downstream Left (DL),. Adapted from [16].

Area

Area of an 'i^{th}' connected component is calculated as:

$$Area_i = \sum_{j=1}^{n} I_i$$

where n is the total number of pixels labelled as i in the ith region or component.

Perimeter

The perimeter of an object is calculated as the total number of pixels on the boundary or edge of that object.

Shape Index(SI)

Shape Index (SI) is calculated as:

$$SI = 4\pi \times \left(\tfrac{Area}{Perimeter^2} \right)$$

4.3. Comparative Analysis Results

Each branch of an artery was subdivided into eight vascular regions, such as Upstream Right (UR), Upstream (U), Upstream Left (UL), Right (R), Left (L),

4. Comparative Analysis of In-vivo Endothelial Cell Data

Downstream Right (DR), Downstream (D), Downstream Left (DL), as shown in Figure 4.3 for cell morphology quantification. Therefore, for all 56 vascular regions examined in all three rabbits (one mature and two immature), 1496 cells and their respective nuclei are analysed. The results obtained are analysed in two different ways:

1. All(1496) cells and their nuclei without separating them according to their ages,

2. *Mature* (559) and *immature* (937) cells and their nuclei separately to see the relationship and changes among them.

	EC			ECn			r^2
	Mean	**SEM/ Branch**	**SEM/ Cell**	**Mean**	**SEM/ Branch**	**SEM/ Cell**	**EC: ECn**
L:W Ratio	7.29	0.08	0.48	2.96	0.02	0.11	0.34
Shape Index	0.25	0.00	0.02	0.71	0.00	0.02	0.32
Angle(^{o})	20.81	1.79	10.28	18.11	1.81	10.40	0.81
Length(μm)	56.84	0.54	3.11	18.76	0.08	0.45	0.21
Width(μm)	8.34	0.06	0.32	6.51	0.02	0.13	0.24
Area(Pixels^2)	294.11	2.77	15.89	92.31	0.38	2.15	0.12
Perimeter(Pixels)	130.05	1.14	6.55	40.89	0.14	0.78	0.17

Table 4.3.: Mean ± SEM per cell and per branch of EC and ECn morphology parameters and their correlation values found for all cells, here total number of cells are 1496.

The axis length, axis width, elongation (i.e. L: W ratios), angle of orientation, circularity (shape index), perimeter, and area of ECn and ECs were calculated automatically as described previously in the Section 4.2.2. The calculations of these parameters are reproducible, and the data are expressed as the mean ± SD over branches and over cells and are calculated automatically using the *Matlab mean*() and *std*() functions unless otherwise mentioned. Histograms and correlation plots between ECs and their ECn, r^2 in all cells and age distributions are given in Figures 4.4, A.1.

78

r^2	L:W Ratio	Shape Index	Orientation	Length	Width	Area	Perimeter
UL	0.24	0.10	0.39	0.83	0.07	0.17	0.22
U	0.19	0.15	0.31	0.87	0.08	0.12	0.22
UR	0.35	0.22	0.44	0.51	0.13	0.31	0.44
L	0.27	0.42	0.38	0.77	0.17	0.24	0.44
R	0.14	0.22	0.25	0.62	0.13	0.12	0.27
DL	0.05	0.44	0.25	0.75	0.18	0.04	0.31
D	0.34	0.16	0.48	0.78	0.10	0.30	0.41
DR	0.18	0.20	0.18	0.79	0.22	0.20	0.12
r^2 **ranges**	0.05-0.35	0.10-0.44	0.18-0.48	0.51-0.87	0.07-0.22	0.04-0.31	0.12-0.44

Table 4.4.: r^2 values of different parameters in different regions in all cells, here total number of cells are 1496.

	EC			ECn			r^2
	Mean	SEM/ Branch	SEM/ Cell	Mean	SEM/ Branch	SEM/ Cell	EC: ECn
L:W Ratio	6.10	0.08	0.46	2.62	0.02	0.10	0.24
Shape Index	0.30	0.00	0.02	0.76	0.00	0.02	0.24
Angle(^{o})	4.82	2.40	13.77	2.65	2.41	13.79	0.73
Length(μm)	50.96	0.56	3.19	17.40	0.08	0.43	0.20
Width(μm)	8.85	0.07	0.40	6.78	0.03	0.15	0.19
Area(Pixels^2)	292.41	3.28	18.80	89.83	0.44	2.51	0.15
Perimeter(Pixels)	118.20	1.18	6.73	38.65	0.13	0.76	0.16

Table 4.5.: Mean $\pm$ SEM over cell and over branch of EC and ECn morphology parameters and their correlation values found for immature cells, here the total number of cells is 937.

4.3.1. Statistical Analysis

Overall, 1496 ECs from the thoracic aortas of 3 rabbits are compared with their ECn to determine the correlation between the cell shape and the shape of its nucleus. In all cells as shown in Figure 4.4, there is positive correlation between the ECs and their ECn for all parameters.

r^2	L:W Ratio	Shape Index	Orien- tation	Len- gth	Wid- th	Ar- ea	Peri- meter
UL	0.10	0.15	0.08	0.84	0.14	0.08	0.04
U	0.19	0.18	0.28	0.83	0.14	0.15	0.24
UR	0.26	0.18	0.32	0.47	0.17	0.20	0.33
L	0.36	0.38	0.40	0.70	0.20	0.29	0.46
R	0.19	0.13	0.14	0.50	0.22	0.19	0.14
DL	0.18	0.35	0.28	0.50	0.27	0.12	0.39
D	0.13	0.08	0.16	0.78	0.10	0.07	0.14
DR	0.08	0.20	0.13	0.69	0.19	0.09	0.11
r^2 **ranges**	0.08-0.36	0.08-0.38	0.08-0.40	0.47-0.84	0.10-0.27	0.07-0.29	0.04-0.46

Table 4.6.: r^2 values of different parameters in different regions in all immature cells, here the total number of cells is 937.

Length:Width Ratio

From Table 4.3, the L:W ratio of ECs is 7.29 ± 0.08 and of ECn is 2.96 ± 0.02 in all regions in all type of animals with different ages. When separating these by age, then in immature cells the L:W ratios of ECs and ECn are 6.10 ± 0.08 and 2.62 ± 0.02, and in mature cells L:W ratios are 9.30 ± 0.15 and 3.55 ± 0.03 respectively. The correlation between ECs and ECn in respect of L:W ratios are 0.34, 0.24, and 0.14 in all cells. In all cells, in all immature and in all mature cells; the mean and SEM of different parameter values are given in Tables 4.3, 4.5, and 4.7 respectively.

Figures 4.5 and A.2 show the vascular region-wise distribution of ECs and ECn L:W ratios in mature and immature animals. A straight line through the origin would represent perfect agreement between cell and nuclear properties. The apparent absence of SEMs for some points arises where the error bar lies within the area of the marker. For the two outliers in the plot of length-to-width ratios, only 14 and 17 cell boundaries could be identified, whereas all other points have total number of cells ≥ 30 [25].

The ranges for L:W ratios for each region in immature cells are [6.13 - 8.14], [2.66 - 3.45] and [0.05 - 0.35], and the ranges found for L:W ratios for each

	EC			ECn			r^2
	Mean	SEM/ Branch	SEM/ Cell	Mean	SEM/ Branch	SEM/ Cell	EC: ECn
L/W Ratio	9.30	0.15	0.83	3.55	0.03	0.17	0.14
Shape Index	0.18	0.00	0.02	0.61	0.00	0.02	0.07
Angle(^{o})	47.98	2.16	12.39	44.36	2.30	13.22	0.89
Length(μm)	66.83	0.99	5.66	21.07	0.11	0.64	0.036
Width(μm)	7.47	0.08	0.46	6.06	0.03	0.19	0.18
Area(Pixels^2)	297.01	4.99	28.66	96.54	0.65	3.74	0.10
Perimeter(Pixels)	150.19	2.09	11.99	44.70	0.21	1.20	0.03

Table 4.7.: Mean $\pm$ SEM over cell and over branch of EC and ECn morphology parameters and their correlation values found for mature cells, here the total number of cells is 559.

r^2	L:W Ratio	Shape Index	Orien- tation	Len- gth	Wid- th	Ar- ea	Peri- meter
UL	0.01	0.01	0.24	0.71	0.00	0.00	0.10
U	0.02	0.07	0.14	0.88	0.10	0.04	0.02
UR	0.13	0.03	0.16	0.82	0.04	0.12	0.08
L	0.09	0.22	0.18	0.07	0.28	0.17	0.04
R	0.01	0.06	0.03	0.64	0.10	0.01	0.00
DL	0.00	0.44	0.17	0.79	0.18	0.00	0.16
D	0.30	0.22	0.39	0.78	0.11	0.26	0.48
DR	0.07	0.39	0.32	0.64	0.04	0.01	0.03
r^2 **ranges**	0.00-0.30	0.01-0.44	0.03-0.39	0.07-0.88	0.00-0.28	0.00-0.26	0.00-0.48

Table 4.8.: r^2 values of different parameters in different regions in all mature, here the total number of cells is 559.

region in all cells are [5.14 - 6.69], [2.37 - 2.81] and [0.00 - 0.30], in mature cells are [7.31 - 11.88], [3.08 - 3.70] and [0.08 - 0.36] respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in mean $\pm$ SEM over branch and SEM over cell, and r^2 values are given in Tables 4.4, 4.6 and 4.8.

It is clear from the plots and table values that the cells elongate more with age, as the difference can be seen more clearly between mature and immature cells' L:W ratios. Also, ECs and ECn parameters are inversely correlated, which demonstrates that more rounded cells and their respective nuclei become more elongated with age, and vice versa. These results are presented in [25].

Orientation

The mean $\pm$ SEM for the angle of orientation of the ECs in all cells is $20.81° \pm 1.79°$, whereas the orientation of ECn is $18.11° \pm 1.81°$. In immature cells, mean $\pm$ SEM the angles are measured $47.98° \pm 2.16°$ and $44.36° \pm 2.30°$, and in mature animals as $4.82° \pm 2.4°$ and $2.41° \pm 13.79°$ respectively, these values can be found in Tables 4.3, 4.5, and 4.7 respectively. The correlation between ECs and ECn in respect to orientation is strongly positive; the values found are 0.81, 0.89, and 0.73 in all cells, in all immature and in all mature cells respectively. The vascular region-wise distribution of EC and ECn angles in mature and immature animals are shown in Figures 4.6 and A.3.

The ranges for means, SEMs and r^2 orientation values for each region in all cells, in immature cells, and in mature are $[-29.91° - 67.10°]$, $[-66.82° - 72.77°]$, $[-41.73° - 67.12°]$, $[-28.06° - 54.11°]$, $[68.46° - 66.42°]$, $[-39.83° - 52.97°]$, $[0.18 - 0.48]$, $[0.03 - 0.39]$ and $[0.08 - 0.40]$ respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in means $\pm$ SEM over branch and SEM over Cell, and r^2 values in Tables 4.4, 4.6, 4.8 in all cells, and in all mature and in all immature cells respectively.

The striking difference between L:W ratios and orientation correlations shows that there is no change in pattern in orientation with age, and also, ECs and ECn are highly aligned, which demonstrates that cells and their respective nuclei tend to become more aligned with blood flow. These results are presented

in [25].

Shape Index

The shape index of ECs is 0.25 ± 0.02, whereas the shape index of ECn is 0.71 ± 0.02. In mature the shape indexes are 0.30 ± 0.02 and 0.18 ± 0.02, and in immature animals shape indexes are measured as 0.76 ± 0.02 and 0.61 ± 0.02 respectively, the mean and SEM values are given in Tables 4.3, 4.5, and 4.7. ECs and ECn shape indexes are positively correlated, the values found are 0.32, 0.24, and 0.07 in all cells, in all immature and in all mature cells. The vascular region-wise distribution of ECs and ECn angles in mature and immature animals are shown in Figures 4.7 and A.4.

The ranges of r^2, means and SEMs for shape index values for each vascular region in all cells are [0.1 - 0.44], [0.21 - 0.30] and [0.63 - 0.76], and the ranges found for shape indexes for each region in immature are [0.01 - 0.44], [0.15 - 0.22] and [0.59 - 0.66], in mature are [0.08 - 0.38], [0.24 - 0.33] and [0.73 - 0.81] respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in means $\pm$ SEM over branch and SEM over cell, and r^2 values in Tables 4.4, 4.6 and 4.8.

Length

In rabbits, the mean cell length is $56.84 \pm 0.54 \mu m$, whereas the mean nuclei length is $18.76 \pm 0.08 \mu m$. In ECs the lengths are $50.96 \pm 0.56 \mu m$ and $66.83 \pm 0.99 \mu m$, and in ECn are $17.40 \pm 0.08 \mu m$ and $44.36 \pm 2.30 \mu m$ for immature and immature animals respectively. The mean and SEM values are given in Tables 4.3, 4.5, and 4.7. The ECn and ECs in length are positively correlated, the values found are 0.21, 0.20, and 0.04 in all cells, in all immature and in all mature cells. The region-wise distribution of ECs and ECn lengths in mature and immature animals are shown in Figures 4.8 and A.5.

The ranges for lengths for each vascular region in mean $\pm$ SEM (ECs, ECn) and r^2 values are $[51.97 \mu m - 63.62 \mu m]$, $[43.15 \mu m - 55.52 \mu m]$ and $[51.09 \mu m - 87.86 \mu m]$, $[17.42 \mu m - 20.52 \mu m]$, $[16.20 \mu m - 18.07 \mu m]$, $[19.02 \mu m - 22.21 \mu m]$, $[0.51 - 0.87]$, $[0.47 - 0.84]$ and $[0.07 - 0.88]$ respectively. Tables 4.5, and 4.7

83

give the detailed results for each region in means $\pm$ SEM over branch and SEM over cell, and r^2 values in Tables 4.4, 4.6 and 4.8.

Width

Mean EC width is $8.34 \pm 0.06\mu m$ while mean ECn width is $6.51 \pm 0.02\mu m$. In immature rabbits, the mean widths are $7.47\mu m \pm 0.08\mu m$ and $6.06\mu m \pm 0.03\mu m$, and in mature animals the widths are measured as $8.85 \pm 0.07\mu m$ and $6.78 \pm 0.02\mu m$ respectively. The mean and SEM values are given in Tables 4.3, 4.5, and 4.7. ECs and ECn widths are positively but very weakly correlated; the values found are 0.24, 0.19, and 0.18 in all cells, in all immature and in all mature cells. The vascular region-wise distribution of EC and ECn angles in mature and immature animals are shown in Figures 4.9 and A.6.

The ranges for widths for each vascular region in immature are $[8.34 - 9.61\mu m]$, $[6.55 - 6.99\mu m]$ and $[0.10 - 0.27]$, and the ranges found for widths for each region in all cells are $[7.59 - 9.18\mu m]$, $[6.13 - 6.79\mu m]$ and $[0.07 - 0.22]$, in mature are $[6.91 - 8.29\mu m]$, $[5.91 - 6.86\mu m]$ and $[0.00 - 0.28]$ respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in means $\pm$ SEM over branch and SEM over cell, and r^2 values in Tables 4.4, 4.6 and 4.8.

Area

Overall, ECs cover area 294.11 ± 2.77 $pixels^2$ whereas ECn covers area 92.31 ± 0.38 $pixels^2$. In immature cells areas are 292.41 ± 3.28 $pixels^2$ and 89.83 ± 0.44 $pixels^2$, and in mature animals areas are measured as 297.01 ± 4.99 $pixels^2$ and 96.54 ± 0.65 $pixels^2$ respectively, the mean and SEM values are given in Tables 4.3, 4.5, and 4.7. There is not a big change in ECs and ECn areas, even with change in age. This demonstrates that, even with change in age, cells and their nuclei do not undergo change in their areas. The correlation between ECs and ECn in area are positive weakly correlated, and the values are 0.12, 0.15, and 0.10 in all cells, in all immature and in all mature cells.

The vascular region-wise distribution of EC and ECn areas in mature and immature animals are shown in Figures 4.10 and A.7. The ranges for areas for

each region in all cells are $[240.19 - 330.63 \ pixels^2]$, $[88.55 - 95.73 \ pixels^2]$ and [0.04 - 0.31], and the ranges found for areas for each region in immature all cells are $[248.54 - 326.60 \ pixels^2]$, $[85.90 - 93.41 \ pixels^2]$ and [0.07 - 0.29], in mature are $[205.67 - 415.71 \ pixels^2]$, $[88.63 - 108.12 \ pixels^2]$ and [0.00 - 0.26] respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in means $\pm$ SEM over branch and SEM over cell, and r^2 values in Tables 4.4, 4.6 and 4.8.

Perimeter

The perimeter of the EC is $130.05 \pm 1.14 \ pixels$ and ECn is 40.89 ± 0.14 *pixels*. In EC, perimeters for immature and mature are $118.20 \pm 1.18 \ pixels$ and $150.19 \pm 2.09 \ pixels$, and in ECn, perimeters are measured as 38.65 ± 0.13 *pixels* and $44.70 \pm 0.21 \ pixels$ respectively, the mean and SEM values are given in Tables 4.3, 4.5, and 4.7. The correlation between ECs and ECn in perimeter is positively correlated, and the results are as 0.17, 0.16, and 0.03 in all cells, in all immature and in all mature cells. The vascular region-wise distribution of EC and ECn perimeter in mature and immature animals are shown in Figures 4.11 and A.8.

The ranges for perimeters for each region in all cells are [120.29 - 143.82 pixels], [38.56 - 43.76 pixels] and [0.12 - 0.44 pixels], and the ranges found for perimeters for each region in immature cells are [102.38 - 127.05 pixels], [36.77 - 39.56 pixels] and [0.04 - 0.46], in mature are [119.44 - 187.94 pixels], [41.12 - 46.76 pixels] and [0.00 - 0.48] respectively in ECs and ECn. Tables 4.5, and 4.7 give the detailed results for each region in means $\pm$ SEM over branch and SEM over cell, and r^2 values in Tables 4.4, 4.6 and 4.8.

4.4. Discussion

In this chapter, additional evidence for similar behaviour of endothelial cells and their nuclei was obtained *in vivo*. Some of the morphological parameters have parallel behaviour which suggests that the mechanical forces caused the same effect in both the cells and their nuclei. Evidence that the time-averaged

magnitude and direction of haemodynamic wall shear stress is likely to be the dominant influence on the morphology of the whole cell has been reviewed in [16]; shear has a greater effect than cyclic strain [194] and mean shear dominates over unsteady components [73].

Considering first length-to-width ratios, the most striking features are low values in immature aortas, and high values in mature aortas. The data are consistent with those previously obtained by Al-Musawi *et al.* [16]. However, the relation is less strong for length-to-width ratios than for alignment, so other influences on elongation, such as pulsatility or wall stiffness (which varies around branches and changes with age), although not considered further, cannot be entirely dismissed [25].

Orientations around immature and mature branches are broadly similar, suggesting that the pattern of flow direction near the wall stayed approximately constant with age. Patterns of elongation and, by implication, near wall flow did change with age in the descending aorta as a whole. The general tendency for nuclear elongation to be greater in mature animals has been extensively discussed elsewhere [26]. The correlations between nuclear and cell orientations around immature and mature branches are highly positive, and in [180] it has been suggested that high lesion frequency correlates with greater nuclear elongation, consistent with lipid deposition occurring in regions experiencing high wall shear stress at both ages. The relevance to human lesions, which also switch from a downstream to a lateral or upstream location around intercostal branch ostia with age [180], remains to be determined.
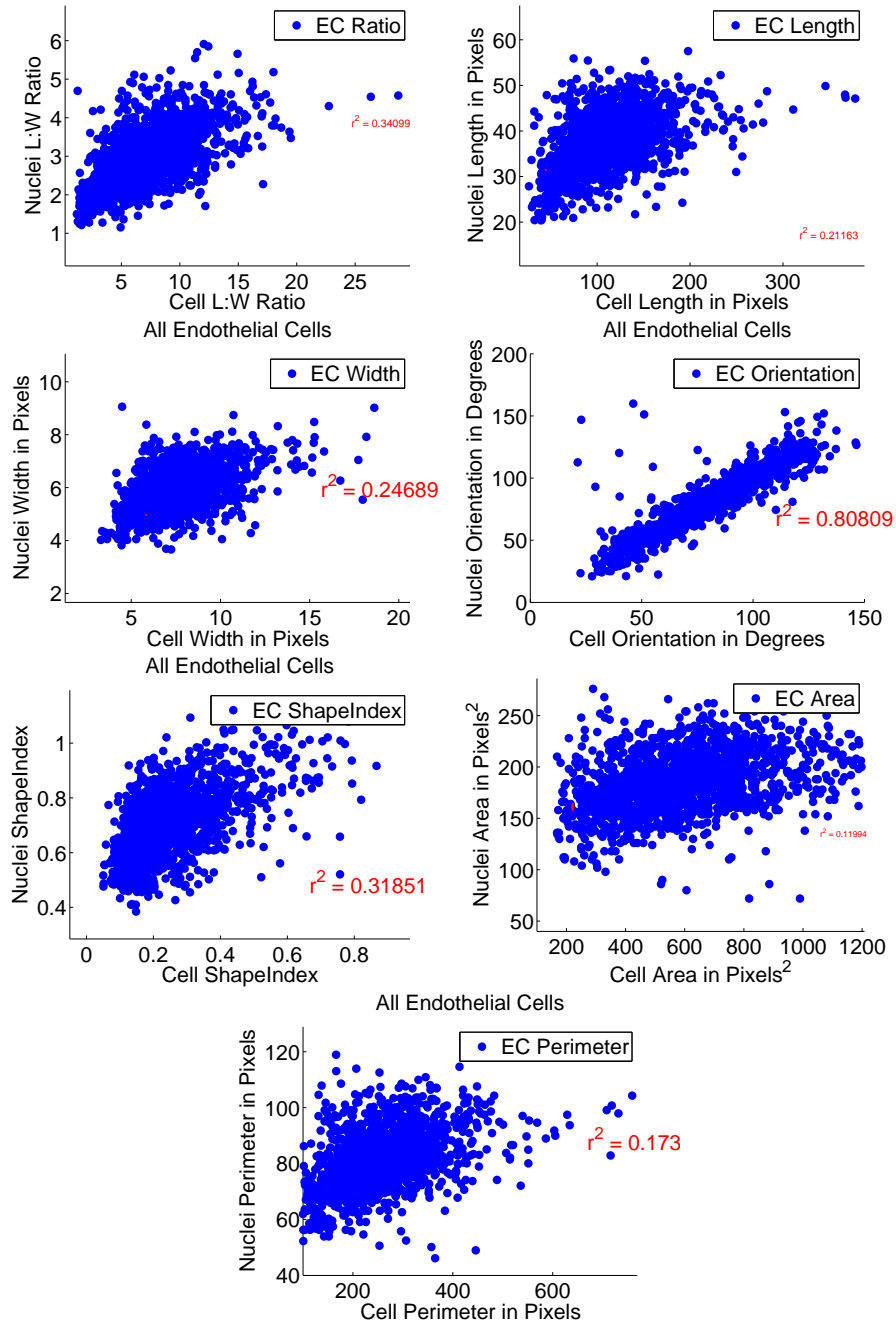
Figure 4.4.: Correlation between cells and nuclei for different shape parameters for all endothelial cells and nuclei, here the total number of cells are 1496.
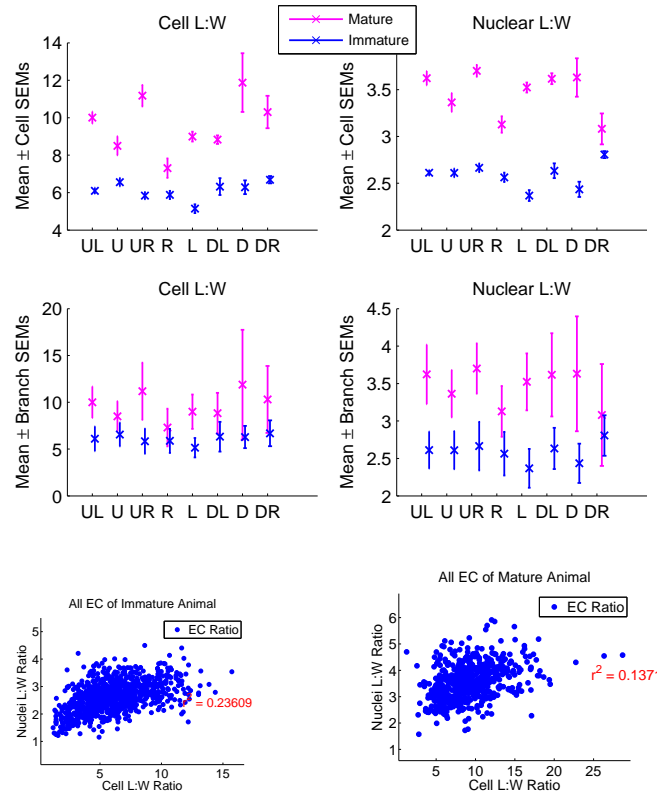
Figure 4.5.: Vascular region-wise ratios and correlation between cells and nuclei ratios, for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
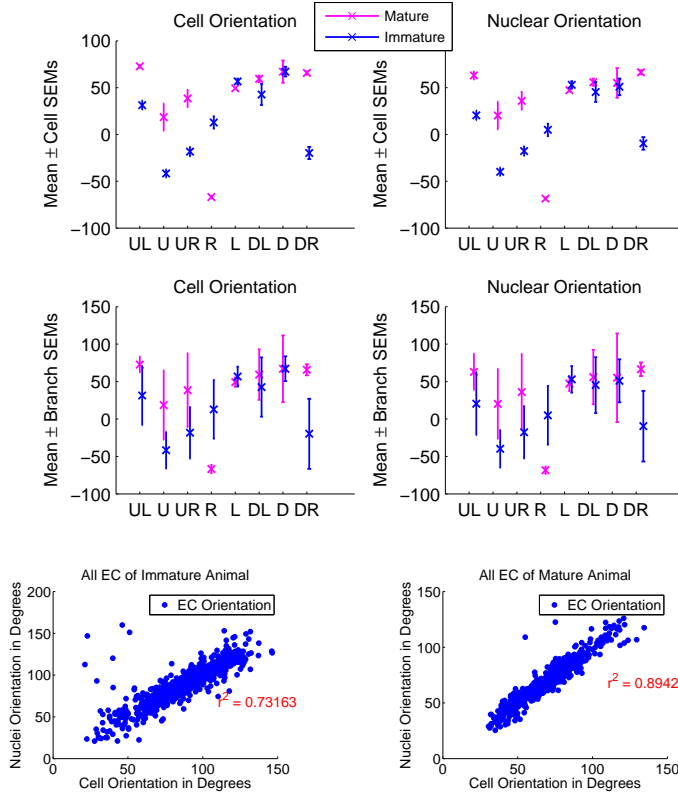
Figure 4.6.: Vascular region-wise and correlation between cells and nuclei orientations for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
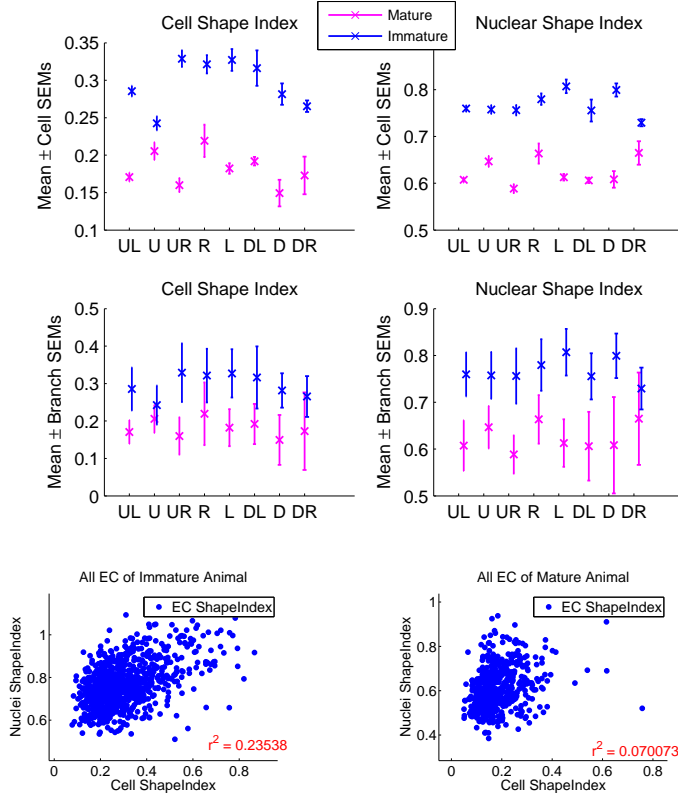
Figure 4.7.: Vascular region-wise shape index and correlation between cells and nuclei for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
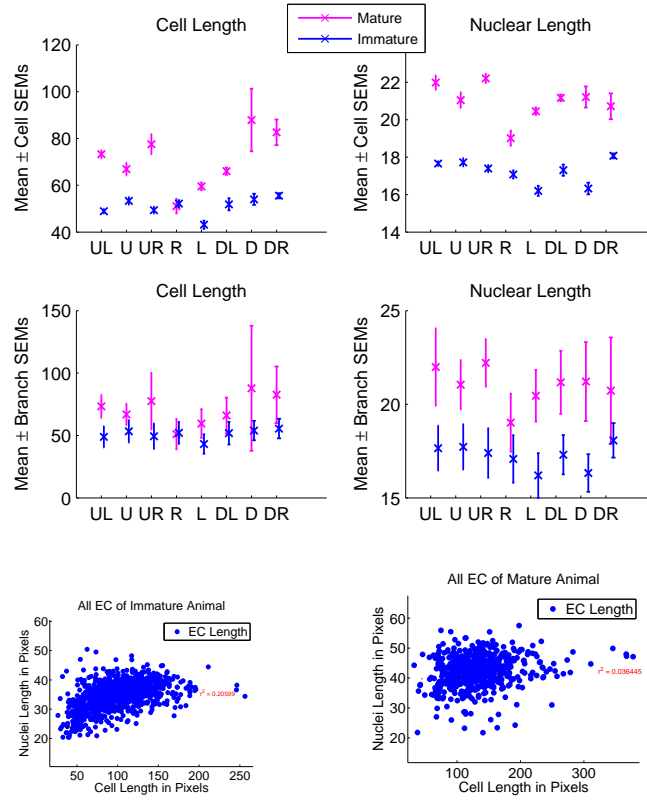
Figure 4.8.: Vascular region-wise lengths and correlation between cells and nuclei for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
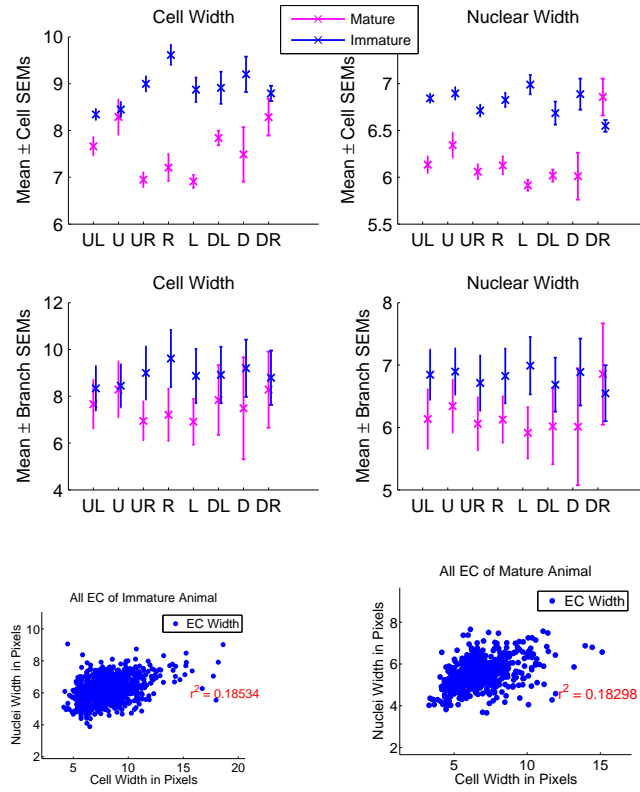
Figure 4.9.: Vascular region-wise widths and correlation between cells and nuclei for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
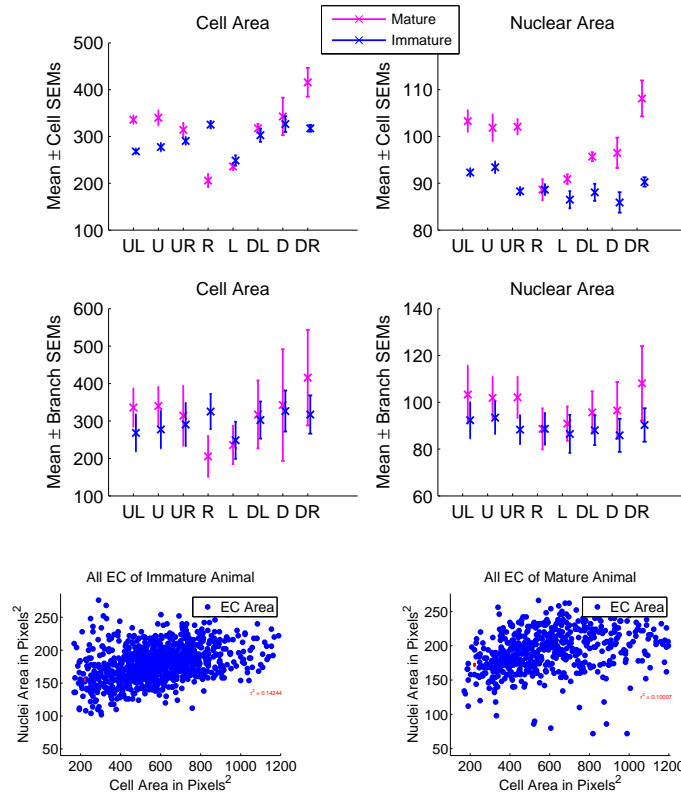
Figure 4.10.: Vascular region-wise areas and correlation between cells and nuclei for mature and immature cells. Here the total number of cells are 559 and 937 respectively.
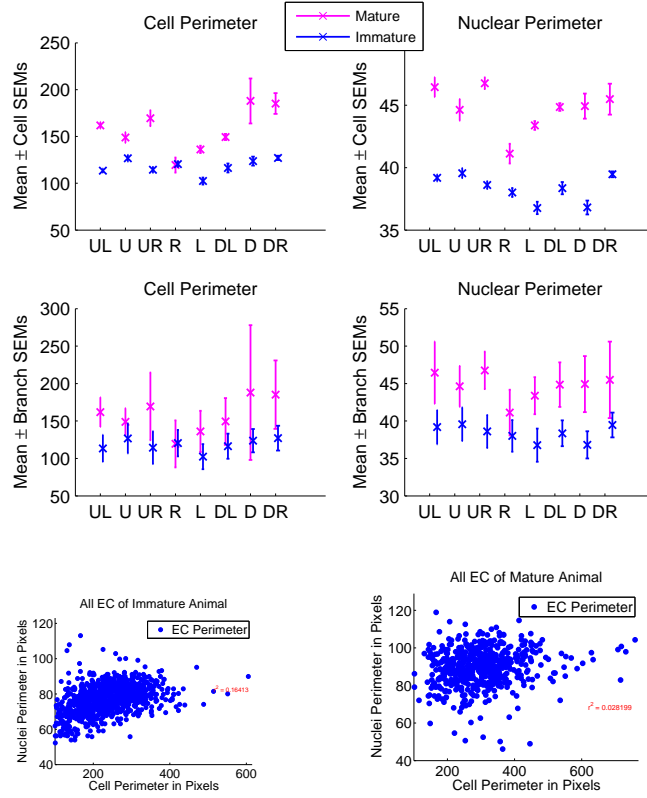
Figure 4.11.: Vascular region-wise perimeters and correlation between cells and nuclei for mature and immature cells. Here the total number of cells are 559 and 937 respectively.

5. Shape Features Detection in *In-vivo* Endothelial Cell Data

5.1. Introduction

From observing the ECs and ECn data, it became clear that there is large variability in cell shapes, even within a single region. In the literature [16, 25, 38, 111], it has been common to study a few early defined parameters describing cell shape and organisation, such as the L: W ratio (ratio between major and minor axis of the shape, which depends on shear stress effects on the cells) and angle of orientation (which depends on the direction of flow). Instead of looking at L: W ratios, and angles of orientations, are there some other properties of cell shape which could be used to annotate or characterise patterns? There is similarity between cells in different regions but not necessarily in similar types of regions. For example, cells of left, right, upstream, or downstream regions of animal 1 might be similar to other regions in the same animal or in different animals, but not necessarily within the same anatomical location.

Therefore, it could be useful to identify different patterns (shapes) in all regions and then see how these might be used to categorise the cell shapes. With this idea in mind, 1496 cells and 7077 nuclei were studied to identify common patterns or shape behaviour in the population.

5.2. Shape Analysis

Shape analysis is an intermediate step in image processing, and usually starts with the representation of the objects using shape descriptors or features.

As shown in Figure 5.1, shape analysis begins with the extraction of features

95

Figure 5.1.: General flow chart of different stages in shape analysis.

from the shape or binarised object. The normalised features are used as an input to some suitable classification algorithm to perform recognition of the shapes or objects. These fundamental stages in shape analysis are discussed in more detail in the following sections.

5.3. Shape Features

To define the shape of an object in a meaningful way, one approach might be to define descriptors or features of the shapes. A shape descriptor should have some robust properties for recognition. Shape descriptors might be defined as that information about an object which is not influenced by size, location, or orientation. In other words, a shape descriptor should be translation, rotation, and scale invariant, and it should form a complete, congruent, and compact set [129].

There are many approaches to shape representation; following [40], these techniques can be broadly categorised into contour-based, region-based and transform-based. The contour-based approach can be further sub-classified into a parametric description of contours, in which the shape outline is represented as a parametric curve; a set of contour points, or the shape outline. Boundary-based shape descriptors use only the object boundary in the description of the object shape. Region-based methods consider the whole area of an object. The region-based approaches can be sub-classified as: region decomposition; bounding regions; and internal features. Transform approaches may be linear or nonlinear.

Shape is a multi-dimensional component of variation in morphological form [61]. The description of patterns of variation in morphological shape is therefore an important step in the recognition of each object and the identification of

96

groups of sub-populations that share morphological features of shape [43]. Some shape analysis applications in biological data are presented in the next Section.

5.4. Shape Features in Biological Images

Shape analysis is an important method in biology to separate broadly similar animals and plants into different groups, by comparing shapes of the specimens. It has evolved into the area of morphometrics, and many methods have been developed to record the positions of "landmarks" on the specimens in 2- or 3-**d** [190]. The basic idea of biological shape analysis is to find variations in objects to understand the physiological changes within those objects. Changes in cellular morphology are important signs in the physiological states of cells, and play an important role in understanding diseases, such as cancer [28] or, in our case, the development of atherosclerotic lesions.

Initially, traditional shape variation approaches also known as traditional morphometrics were focused on to use a linear combination of the inter-landmark distances in biological structures, as well as angles and distance ratios. Most modern geometric morphometric methods, which define shape geometrically, have replaced these traditional morphometic approaches [125]. Statistical shape analysis is usually based on "Kendall's shape space" [93, 94].

Some applications of shape analysis for biological problems have been previously reported. Different moments are applied as feature descriptors, such as moment invariants [78], and Zernike moments [161]. By using artificial neural networks (ANNs) as a classifier and image processing techniques, Yang *et al.* [189] developed an automatic routine examination system for parasitic diseases. For the classification of closed planar shapes using a neural network is presented in [66]. Chain codes [57] and Fourier descriptors [137] are the most commonly used boundary shape descriptors. For detecting morphological variability of Eimeria species, Sommer [157] developed a parametric contour approach using the Fourier transform. Foran *et al.* in [55] developed an image retrieval system which uses textural and similarity invariant Fourier descriptors for shape characterisation to discriminate between malignant lymphomas and chronic lymphocytic leukaemia. Multi-scale mathematical morphology has been

applied for automatically extract features of contours to diatom classification by using decision trees and k-nearest neighbour in [87]. For the characterisation of cat ganglion neural cells used multi-scale features, whereas for the same problem Costa *et al.* [44] addressed by proposing a comprehensive set of shape measures, such as eccentricity, fractal dimension, influence area, convex hull area and with the help of several clustering strategies. To understand cell motility, several techniques have been proposed such as in [64]. Cell shape analysis methods should not add spurious information, and should not discard important information [140]. They should also capture biologically important shape variations in cells [115]. A comparison of quantitative methods for cell-shape analysis is presented in [140], while [36, 115, 135] provide comprehensive reviews for shape analysis methods.

5.5. Morphological and Statistical Features

The selection of the features has been identified as the most critical phase in shape recognition, even more difficult than the classification step. As suggested in the previous discussion, the description and measurement of biological shape description and measurement is a fundamental step to assess variation in natural populations and also to infer its ecological and evolutionary causes [43]. ECs can be subjected to many morphological analyses. The motivation of this study is to provide an appropriate descriptive technique for vascular ECs and ECn shape which could relate to some relevant biological properties of the cells and nuclei.

ECs and ECn shapes can be mathematically understood as a set of connected points in a 2-dimensional image, which could be transformed into **d**-dimensional feature space Φ. As explained earlier, the transformation space, Φ, is usually translation, rotation and scaling invariant.

In the present shape study two transformation methods are proposed: one, shape-descriptor analysis, in which several geometric shape features are extracted, and another shape-boundary contour analysis. These techniques are explained below:

5.6. Shape Feature Descriptor

The features selected for ECs and ECn include:

1. Compactness, which is the indicator of the circularity of the object, and is given by:

$$C = 4\pi \left[\frac{A}{P^2} \right]$$

   Where A is area and P is perimeter.

2. Length : Width ratio, which is given by:

$$LW\,Ratio = \frac{L}{W}$$

   Where L is length and W is width.

3. Eccentricity, calculated as:

$$E = \left[\frac{\sqrt{L^2 - W^2}}{1/2(L)} \right]$$

4. Ratio of the difference between maximum and minimum distance (R) with major and minor axis

$$AR_{max} = \frac{[\arg\max(R) - \arg\min(R)]}{(L-W)}.$$

5. Ratio between length and perimeter

$$PerLengthR = \frac{L}{P}.$$

6. Percentage of number of boundary points within the ideal ellipse. The ideal ellipse can be defined as "A closed, symmetric curve shaped like an oval, which can be formed by intersecting a cone with a plane that is not parallel or perpendicular to the cone's base. The sum of the distances of any point on an ellipse from two fixed points (called the foci) remains constant no matter where the point is on the curve" [10]. The ideal ellipse for one nucleus found using its angle of orientation and major, and minor-axes and its corresponding perimeter points are plotted in Figure 5.2.

99

$$B_{in} = \frac{\sum_{i=1}^{n} P_i}{n} \times 100$$

if a boundary point is inside the ellipse then P_i is one, else P_i is zero.

7. Fourier descriptors (FD) [137] of the signals.
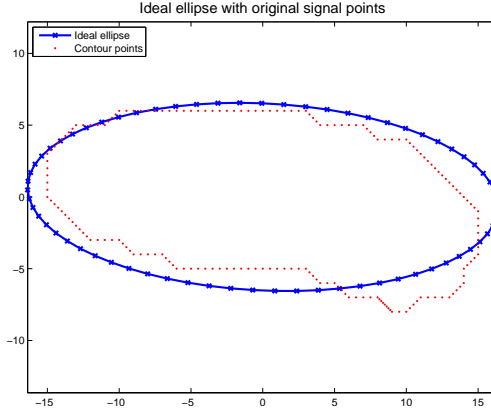
Figure 5.2.: Ideal ellipse (in blue), and original nucleus with boundary points (in red).

The feature vector of each cell boundary and nucleus selected for further shape analysis is then constructed as:

$$\mathbf{SD} = [C, LW Ratio, E, AR_{max}, PerLengthR, B_{in}, FD]. \qquad (5.1)$$

Some of these shape descriptors have been widely applied in object recognition. Each descriptor alone is insufficient for a complex recognition task, but their combination has shown good recognition capabilities [120]. All shape descriptors are normalised with a zero mean and a unit variance, as given in Equation (2.2), before further analysis.

5.7. Contour-based Analysis

After the registration operation as explained in Section 4.2, a set of selected cells and their respective nuclei regions were found. Selected cell contours on
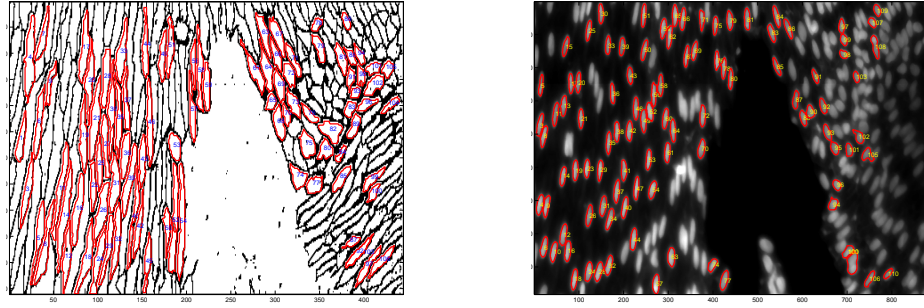
Figure 5.3.: One sample pair of nuclei and cells images with trace boundaries overlaid in red after applying the algorithm 4.1 on the Figure 3.19.

one image set were overlaid, as shown in Figure 5.3.
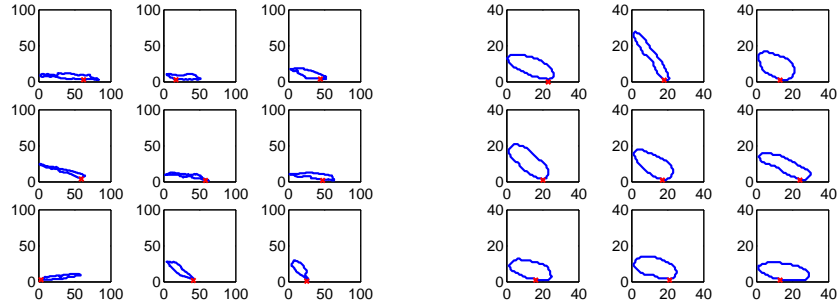
Figure 5.4.: A few samples of original cells and nuclei contours. All axes are calibrated in pixels.

The curvature of each object (cell or nucleus) was calculated using its x and y locations of the edges as a $1\mathbf{d}$ complex signal, or the perimeter contour points of each object, where (x, y) locations of objects are expressed as complex numbers in $(x + iy)$ form. A few samples of nuclei and cell contours are shown in Figure 5.4.

A smoothing operation was then applied, by an interpolation given in Equation (3.3) in Chapter 3, to all the cells and nuclei to reduce sampling artefacts in the perimeter contours. A few samples of the smoothed nuclei and cell per-

101
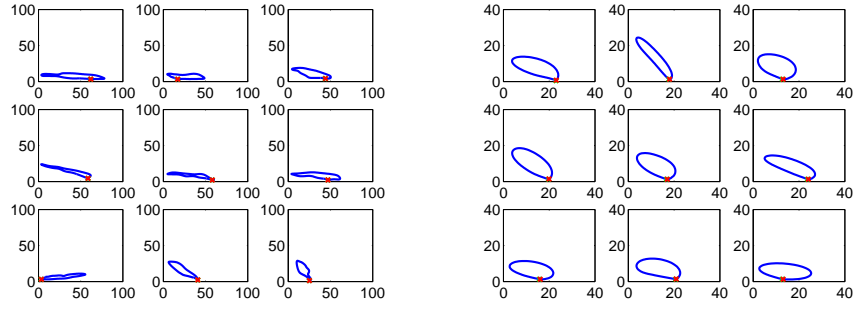
Figure 5.5.: A few samples of smoothed cells and nuclei contours. All axes are
calibrated in pixels.

imeter contours are shown in Figure 5.5.
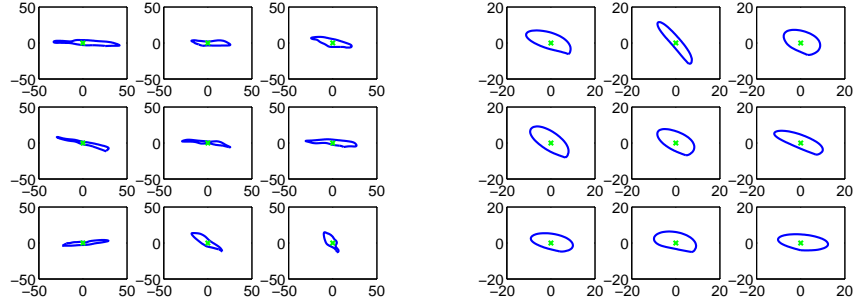
Figure 5.6.: A few samples of centered cells and nuclei contours. All axes are
calibrated in pixels.

All cells and nuclei contours or signals were then aligned to their respective
centroids. Samples of the centred nuclei and cell contours are shown in Figure
5.6.

To achieve rotation invariance, the cell and nuclei contours were rotated with
respect to the zero-degree as shown in Figure 5.7. The starting point problem
in contours can be addressed in many biological contexts by finding the longest
diameter of the shape. By taking the point which was nearest to zero-degrees;
the object can be shifted left or right along the contour with a suitable angle
of reference to make almost all objects to start from the same point. A few
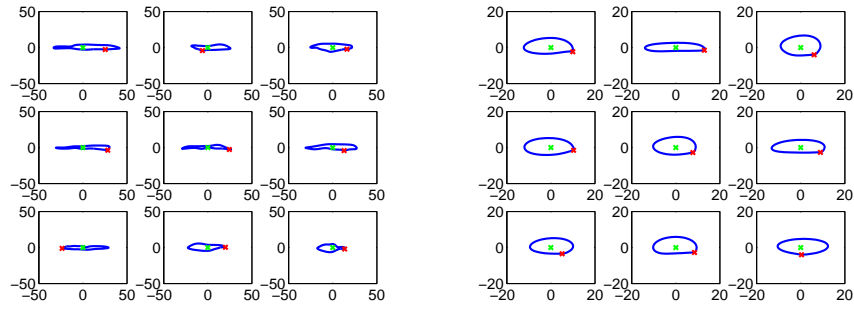
Figure 5.7.: A few samples of cells and nuclei contours after being rotated to zero degree. All axes are calibrated in pixels.

samples of shifted, centred, translated and rotated cells and nuclei are shown in Figure 5.8. The reconstructed averages of cells and nuclei obtained from all the perimeter contours available are given in Figure 5.9.
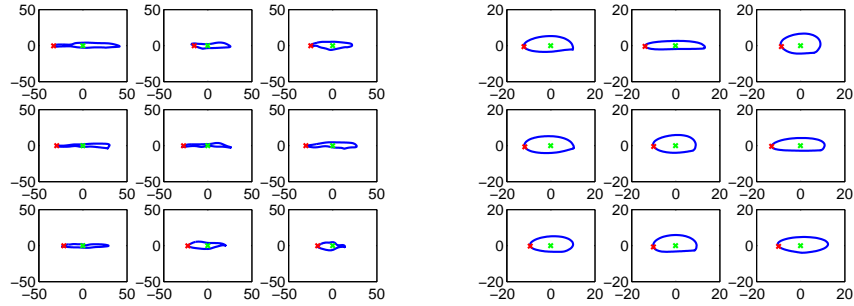
Figure 5.8.: A few samples of centred and shifted cells and nuclei contours. All axes are calibrated in pixels.

5.8. Clustering Results

An unsupervised hierarchical clustering technique was adopted to cluster morphological appearances in the cells and in the nuclei. This was implemented using the functions *dendrogram*(), *linkage*() with parameters *weighted, complete*
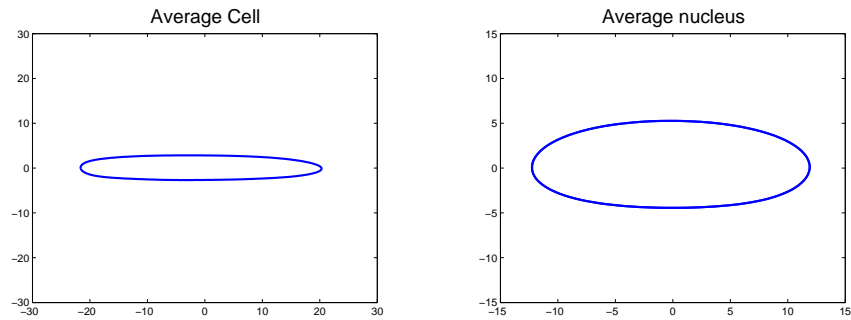
Figure 5.9.: "Average" cell estimated from 1496 cells, and "average" nucleus estimated from 7077 nuclei respectively.

and average, and *pdist*() function with *Euclidean, and cosine* distances, given in the statistical toolbox of *Matlab* version 7.9.0 [121] on a PC running Windows.

1. The distances between each object were calculated first. The *pdist*() function calculates the distances between feature vectors of every pair of cells or nuclei in the data, resulting in a square distance matrix of size equal to the data.

2. Then on the resulting distance matrix, or a more general dissimilarity matrix found through step (1), an agglomerative hierarchical cluster tree was generated with the help of the *linkage*() function.

3. Finally, a dendrogram plot was generated with the functions *dendrogram*(). They are sort of *intersect* shaped: (⊓) lines connecting objects in a hierarchical tree. The height of each '⊓' represents the distance between the two objects being connected.

The hierarchical trees were obtained with combinations of *cosine* distance, and *weighted*, and *complete* linkages obtained for cells and nuclei with the shape descriptors described in Sections 5.8.2 and 5.8.3.

	ECs		ECn	
Cluster Validation Index	**BCNS**	**BCNSFV**	**BCNS**	**BCNSFV**
Silhouette index	3	2	2	3
Davies-Bouldin	5	4	8	5
Calinski-Harabasz	2	10	2	2
Dunn index	3	3	2	2
Hubert-Levin (C-index)	7	5	9	7
Krzanowski-Lai index	3	5	5	9
weighted inter-intra index	7	7	9	7
Range of Indices	[2-7]	[2-10]	[2-9]	[2-9]

Table 5.1.: Summary of cluster validation indices for ECs and ECn. Here BCNS is Best Cluster Number in Signal, and BCNSFV is Best Cluster number in Shape Feature Vector.

5.8.1. Clustering Validity

Cluster validation was done using the *Matlab* cluster validity toolbox available online [11]. This tool allows us to find the optimum number of clusters in a given data and produces a set of indices based on three categories; external assessment of validity, an internal examination of validity and a relative test, as discussed earlier in the review Chapter 2. Some of the indices this tool gives are: Cluster Validation Index, Silhouette index, Davies-Bouldin, Calinski-Harabasz, Dunn index, Hubert-Levin (C-index), Krzanowski-Lai index and weighted inter-intra index, a detailed description of these indices is given in [17, 19, 86, 122]. In this, different cluster validity indices are calculated for both cells and nuclei data sets (shape descriptors and perimeter contours separately) and these indices have given an optimum number of clusters in each data set. The optimum number of clusters found from each index is shown with a square box in Figures 5.10, 5.11, 5.12, and 5.13 and given in Table 5.1.

From Table 5.1, it can be observed that using the perimeter contours for vascular ECs, the minimum number of clusters could be 2 and the maximum 10. When looking at the shape descriptor cluster results in Table 5.1, there is large variation in the estimated optimal number of clusters. It was therefore hard to determine an ideal number of clusters. Therefore, 8 clusters are selected

in order to preserve as much distinction as possible without "over-partition" the data.

Similar observations were made on nuclei data. In Table 5.1, the ECn cluster validity indices results and show great variation in the optimal number of clusters between methods. So, again for both nuclei data types, 8 clusters were selected for classification of the nuclei data. The clustering results are presented in the next subsections individually for both ECs and ECn; shape descriptors and perimeter contours are documented separately.

5.8.2. Clustering based on Shape Feature Descriptors

For 7077 nuclei, shape-descriptors based on dendrogram trees are computed and given in Figures 5.14(a) and 5.15(a), and the average shapes for each of 8 groups are given in Figures 5.14(b) and 5.15(b). Similarly, for 1496 cell shapes, dendrogram trees are given in Figures 5.16(a) and 5.17(a), and eight average cell shapes are presented in Figures 5.16(b) and 5.17(b) respectively. For each of the 8 clusters, if any group has less than five members, then for this group, the average was not calculated. Those cells or nuclei are displayed on top of each other with dashed curves. To visualize the classification results, the cluster indices are mapped with different colours on the original data (cells or nuclei). For the nuclei, Figures 5.14(c-d) and 5.15(c-d) represent the results of different hierarchical combination indices on two different branches. For cells, a similar mapping was done and the results are displayed in Figures 5.16(c-d) and 5.17(c-d).

5.8.3. Clustering based on Contours

Contour-based dendrogram trees for 7077 nuclei and 1496 cells are also computed and given in Figures 5.18(a), 5.19(a), 5.20(a) and 5.21(a) respectively. Average shapes obtained for all nuclei in their corresponding 8 clusters are given in Figures 5.18(b), 5.19(b), 5.20(b) and 5.21(b) respectively. Again, to visualize the clustering results found using contour-based hierarchical trees, data is then mapped with different colours, as shown in Figures 5.18(c-d), 5.19(c-d), on two different regions for nuclei. For cells, a similar mapping is done and results are

Figure 5.10.: Validation indices found for cell signals; an optimal number of clusters in each index is marked with a square box.

Figure 5.11.: Validation indices found for cell shapes; an optimal number of clusters in each index is marked with a square box.
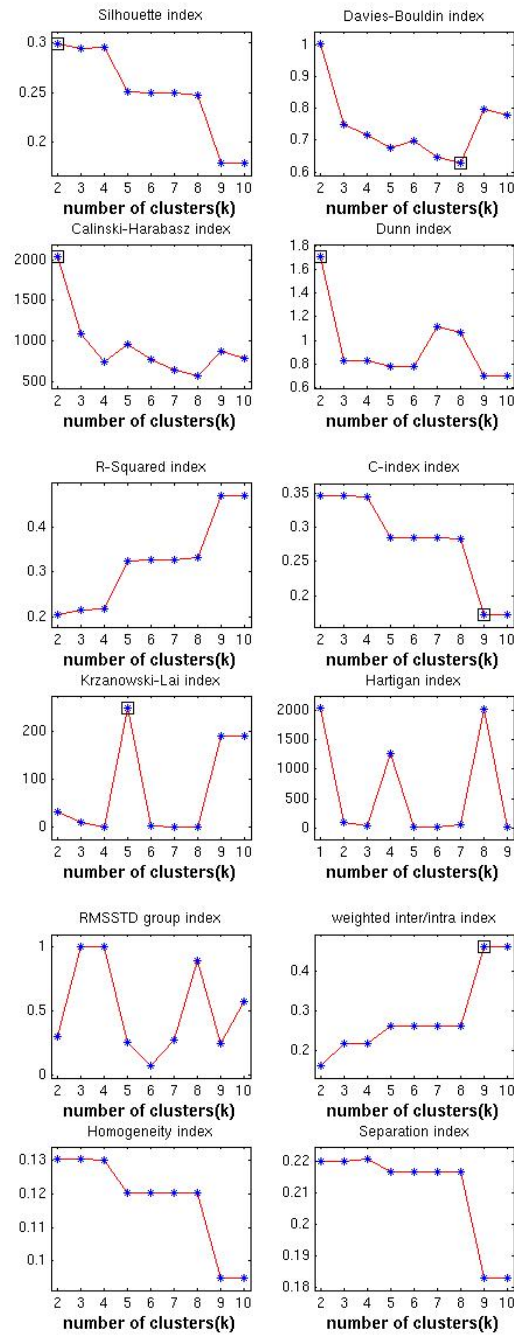
Figure 5.12.: Validation indices found for nuclei signals; an optimal number of clusters in each index is marked with a square box.

Figure 5.13.: Validation indices found for nuclei shapes; an optimal number of clusters in each index is marked with a square box.

Figure 5.14.: A dendrogram showing 8 clusterings using *cosine* distance and *complete* linkage, reconstructed "average" morphology in each of 8 clusters for ECn shape descriptors, and colour mapping examples on two vascular regions of 7077 ECn shape descriptors.

Figure 5.15.: A dendrogram showing 8 clusterings using *cosine* distance and *weighted* linkage, reconstructed "average" morphology in each of 8 clusters for ECn shape descriptors, and colour mapping examples on two vascular regions of 7077 ECn shape descriptors.
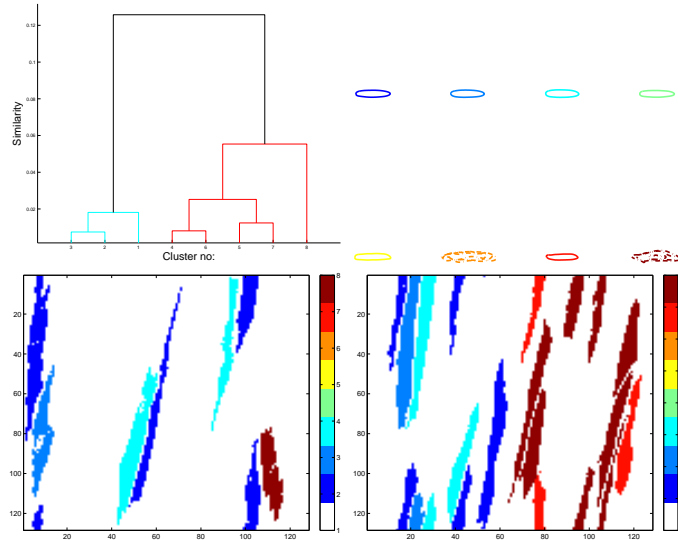
Figure 5.16.: A dendrogram showing 8 clustering using *cosine* distance and *complete* linkage, reconstructed "average" morphology in each of 8 clusters for ECn shape descriptors, and colour mapping examples on two vascular regions of 1496 ECs shape descriptors.

displayed in Figures 5.20(c-d), 5.21(c-d).

5.9. Discussion

While literature on ECs and ECn has concentrated mostly on L:W ratios and angles of orientation, two novel shape analysis techniques are proposed in this chapter: one, shape-descriptor analysis, in which several geometric shape features are extracted, and another, shape-boundary contour analysis. These methods are developed to find modes of shape variation and some other properties beyond just elongation and aspect ratio among the given cell and nuclear data. These techniques lay a basic foundation for further studies in cell-morphology analysis.

Due to time constraints, a detailed biological interpretation has yet to be done to give a firm conclusion on these shape results. But, initial analysis on nuclear data show some variations in shape clusters of immature and mature
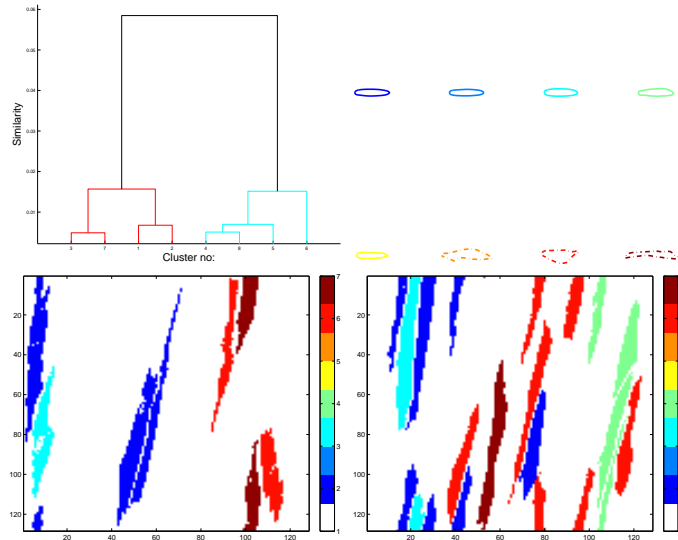
Figure 5.17.: A dendrogram showing 8 clustering using *cosine* distance and *weighted* linkage, reconstructed "average" morphology in each of 8 clusters for ECn shape descriptors, and colour mapping examples on two vascular regions of 1496 ECs shape descriptors.
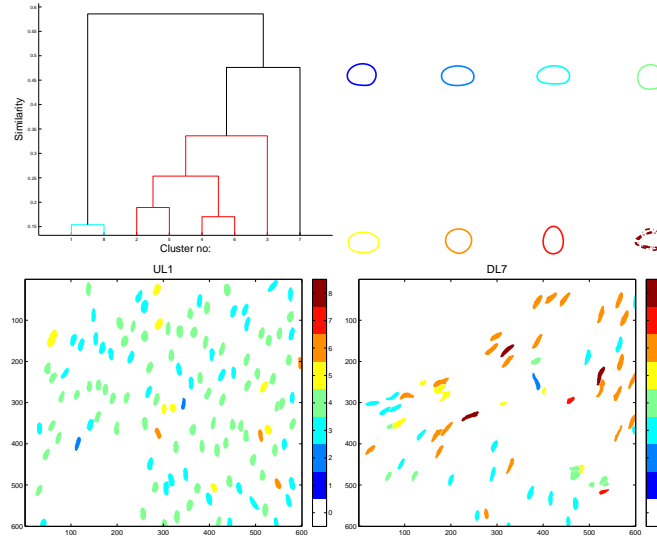
Figure 5.18.: A dendrogram showing 8 clustering using *cosine* distance and *complete* linkage, reconstructed "average" morphology in each of 8 clusters for ECn perimeter contours, and colour mapping examples on two vascular regions of 7077 ECn perimeter contours.
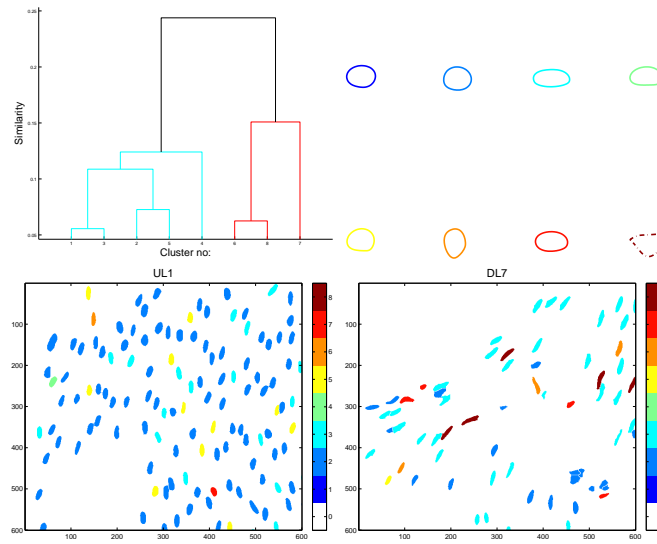
Figure 5.19.: A dendrogram showing 8 clustering using *cosine* distance and *weighted* linkage, reconstructed "average" morphology in each of 8 clusters for ECn perimeter contours, and colour mapping examples on two vascular regions of 7077 ECn perimeter contours.

Figure 5.20.: A dendrogram showing 8 clustering using *cosine* distance and *complete* linkage, reconstructed "average" morphology in each of 8 clusters for ECn perimeter contours, and colour mapping examples on two vascular regions of 1496 ECs perimeter contours.
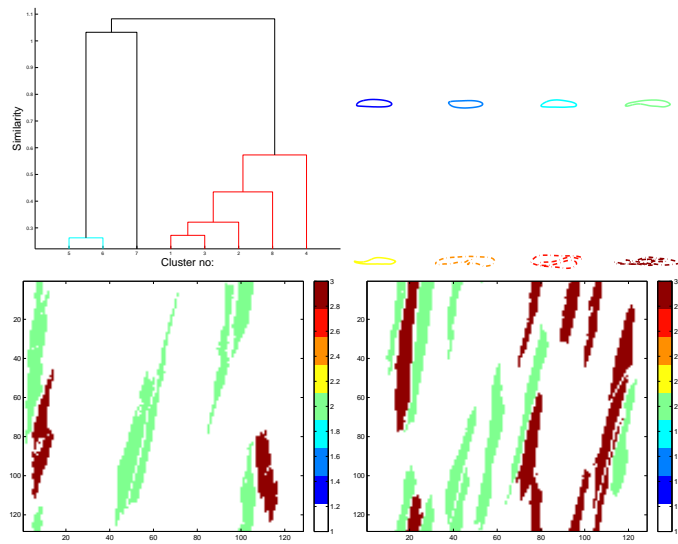
Figure 5.21.: A dendrogram showing 8 clustering using *cosine* distance and *weighted* linkage, reconstructed "average" morphology in each of 8 clusters for ECn perimeter contours, and colour mapping examples on two vascular regions of 1496 ECs perimeter contours.

animal nuclei data, these results are presented in Figures 5.22 and 5.23. In these plots, two distributions (immature and mature animal nuclei) are presented for both shape analysis techniques, i.e. feature descriptors and contour-based methods, displaying a range of cluster indices in each distribution. However, it is clear that the incidence of a few clusters is different in both groups. Initial analysis on nuclear data clearly show variation in shapes of immature and mature animal nuclei data. The representation of these shapes with respect to the atherosclerotic lesion sites and changes in patterns with age are yet to be determined.
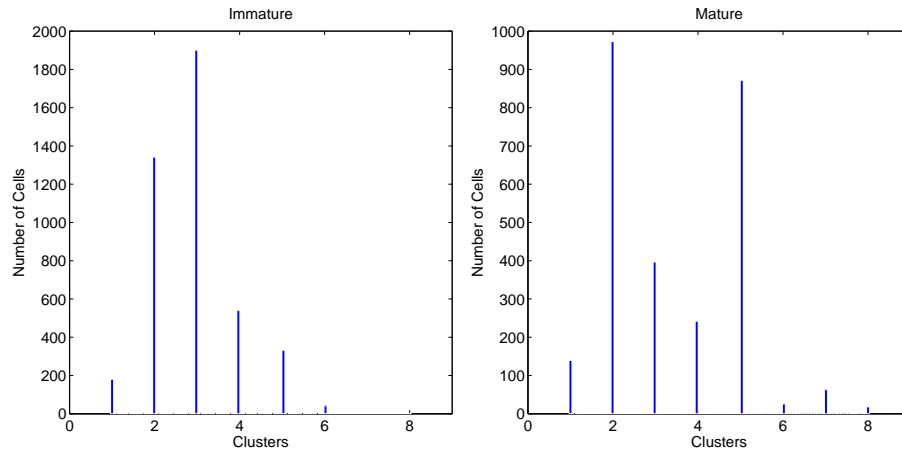
Figure 5.22.: Cluster indices distribution based on perimeter-contour method in all immature and mature nuclei. Here, the total number of nuclei are 4346 and 2731 respectively. Note the incidence of cluster 5 type index in mature rabbit nuclei.
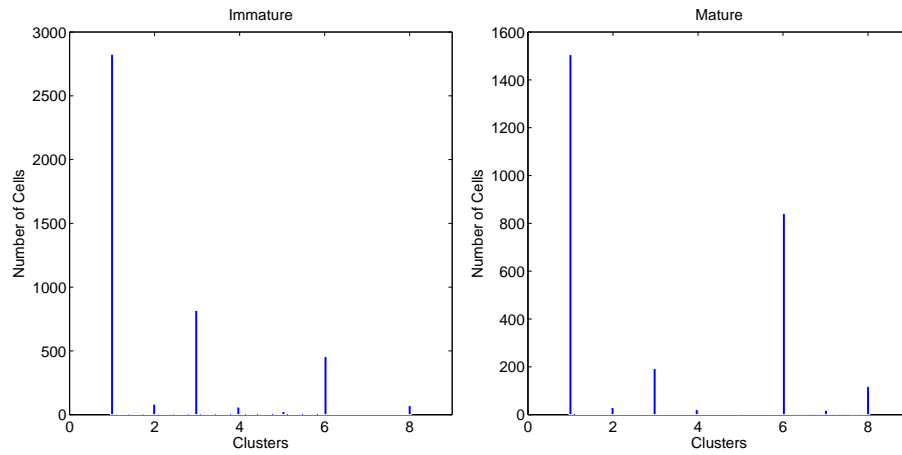
Figure 5.23.: Cluster indices distribution based on shape descriptor method in all immature and mature nuclei. Here, the total number of nuclei are 4346 and 2731 respectively. Note the increased incidence of cluster 6 type index in mature rabbit nuclei.

6. Experiments on *In-vitro* Cells (Temporal Data)

6.1. Introduction

This chapter describes the experiments done, and new analysis methods developed in order to study the role of calcium waves in oscillations under different flow conditions on image sequences of cells in flow field over time. The methods are presented in this chapter to analyse the spatial and temporal behaviour of single-cells amongst their neighbourhood using calcium sensitive fluorescent probes of two different wavelengths, i.e. red and green fluorescence channels. Two different series of images, represented with green and red florescence, sampled at approximately at a frame rate of, $f(5s)$ or 1 frame per 5 second, were acquired as described in Chapter 1.7.2.

6.2. Analysis

6.2.1. Cell Database

These image sequences are treated as stacks. An image of maximum intensity projection over time is found using Equation (6.1):

$$I_{Max}(x, y) = \max_{i=1:N} I_i(x, y), \qquad (6.1)$$

where I_i is the ith frame in one of the green channel stacks, represented in Figure 6.1 and N is the number of frames in the stack. $I_i(x, y)$ is the intensity value at pixel location (x, y), and $i \subseteq 1, \cdots N$.

Instead of taking the first frame of the stack as the frame of reference for
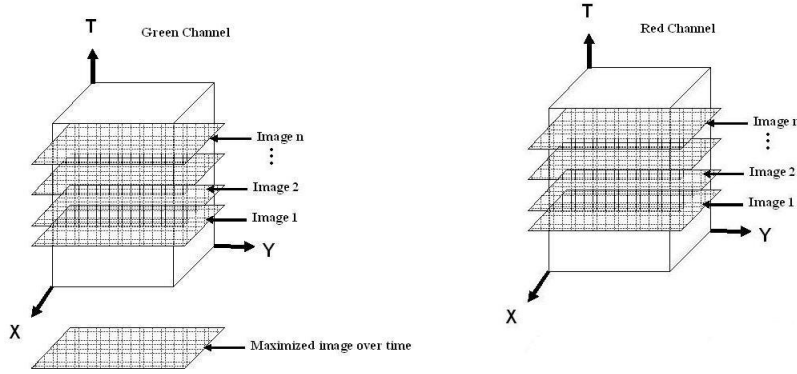
Figure 6.1.: Maximum intensity projection over time image in the green and red channels representation.

finding the binary image or finding binary images of all the frames in the stack, MIPT image is found so that just a single image can be used as a stable frame of reference for each cell. This avoids the need to solve a multi-region and multi-frame tracking problem, though it may lead to errors in processing if a cell moves significantly over the course of the experiment. Equation (6.1) searches over all the N frames in the stack at location (x, y) and finds the maximum intensity value. This process is repeated until it finds the maximum intensities for all the pixels in the stack. This image is termed "maximum intensity projection over time image (MIPT)" and denoted as $I_{Max}(x, y)$, as formulated in Equation (6.1), and shown in Figure 6.2.

A threshold, t_2, is applied to the maximum intensity projection image to find a binary image to use as a mask. Connected component labelling [48] is used to enumerate the regions in the binary image I_{Binary}, given by Equation (6.2) and shown in Figure 6.2.

$$I_{Binary}(x, y) \quad = \quad I_{Max}(x, y) \geq t_2. \tag{6.2}$$

A set of cell regions C, as shown in Figure 6.2, is created using connected

118
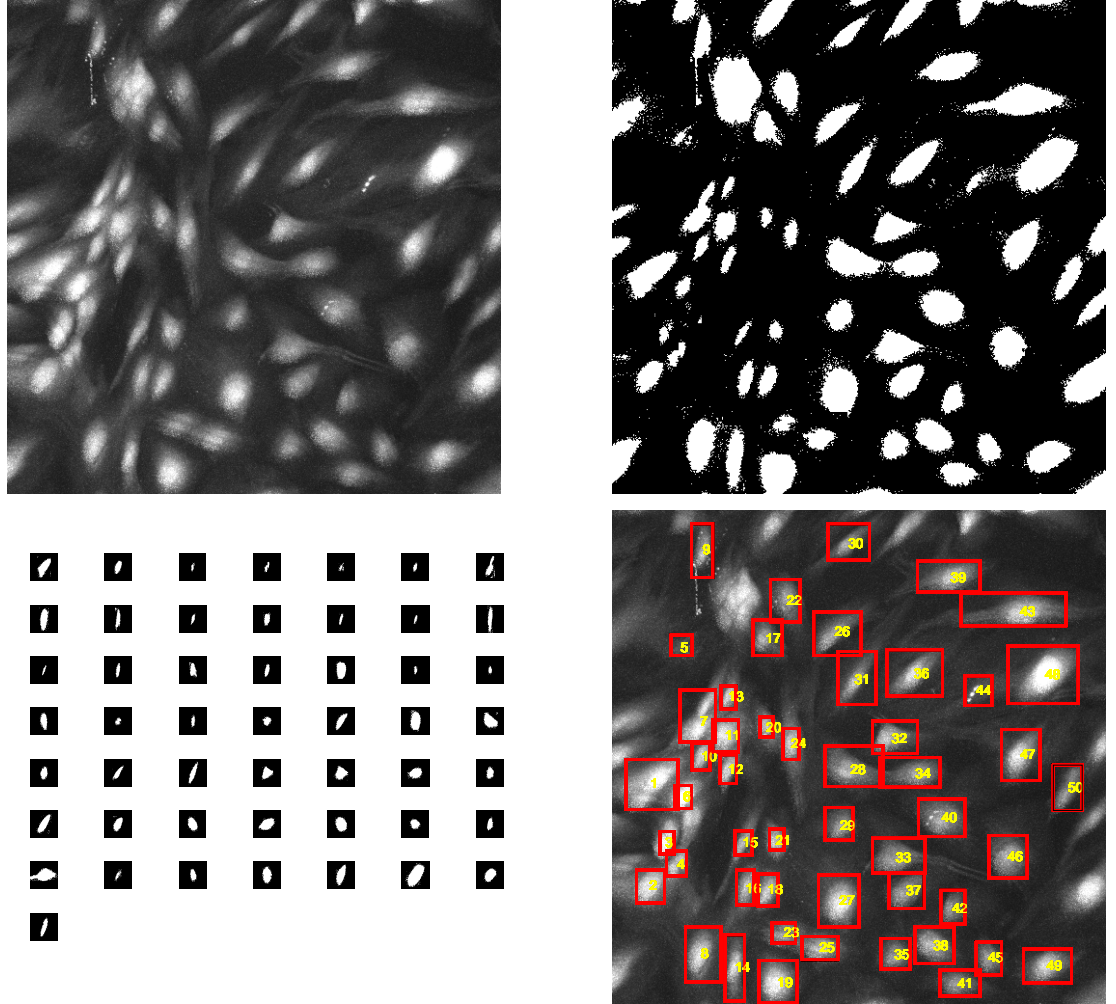
Figure 6.2.: Top panel, original MIPT image, and binary image. Bottom panel, masks of 50 cell regions, and selected regions representation on the MIPT image.

component labelling and the centroid of each region is found using moment calculations. From the selected regions shown in Figure 6.2, the mean intensity is calculated for each cell over all the frames in the red stack:

$$\mu_r(x,y) = \frac{\sum f_{r,i}(x,y) \cdot C_j(x,y)}{N \times \alpha_j}, \quad \forall i \subseteq 1, \cdots N, \quad \forall j = 1 : M, \qquad (6.3)$$

where C_j is the selected j^{th} cell in the ith frame in the stacks, $C_j(x,y)$ takes values of 1 for pixels in the jth region and zero otherwise, α_j is the total number of pixels in the selected cell C_j, N is the total number of frames in the stack, and M is the total number of selected cells. The same process was repeated with the other, green stack to calculate $\mu_2(x,y)$.

$$\mu_g(x,y) = \frac{\sum f_{g,i}(x,y) \cdot C_j(x,y)}{N \times \alpha_j}, \quad \forall i \subseteq 1, \cdots N, \quad \forall j = 1 : M, \qquad (6.4)$$

Finally, the fluorescence ratio, $r_f(x,y)$, given in Equation (6.5), is calculated to estimate the calcium proportionate for each frame in the stack. Here, f_g and f_r are the green and red fluorescent image stacks.

$$r_f(x,y) = \frac{\mu_g(x,y)}{\mu_r(x,y)}. \qquad (6.5)$$

6.3. Results

The calcium fluorescence ratio trend signals, $r_f(x,y)$, in a few example cells during and after flow are shown in Figure 6.3. There are some cells whose Ca^{2+} concentration remains the same, while some cells oscillate more, and in some cells, the value of the $r_f(x,y)$ increases in magnitude. This demonstrates the EC heterogeneity between cells. In order to further analyse the behaviour of endothelial cells with flow, the perfusion sequence was separated into "during flow" and "after flow" conditions.
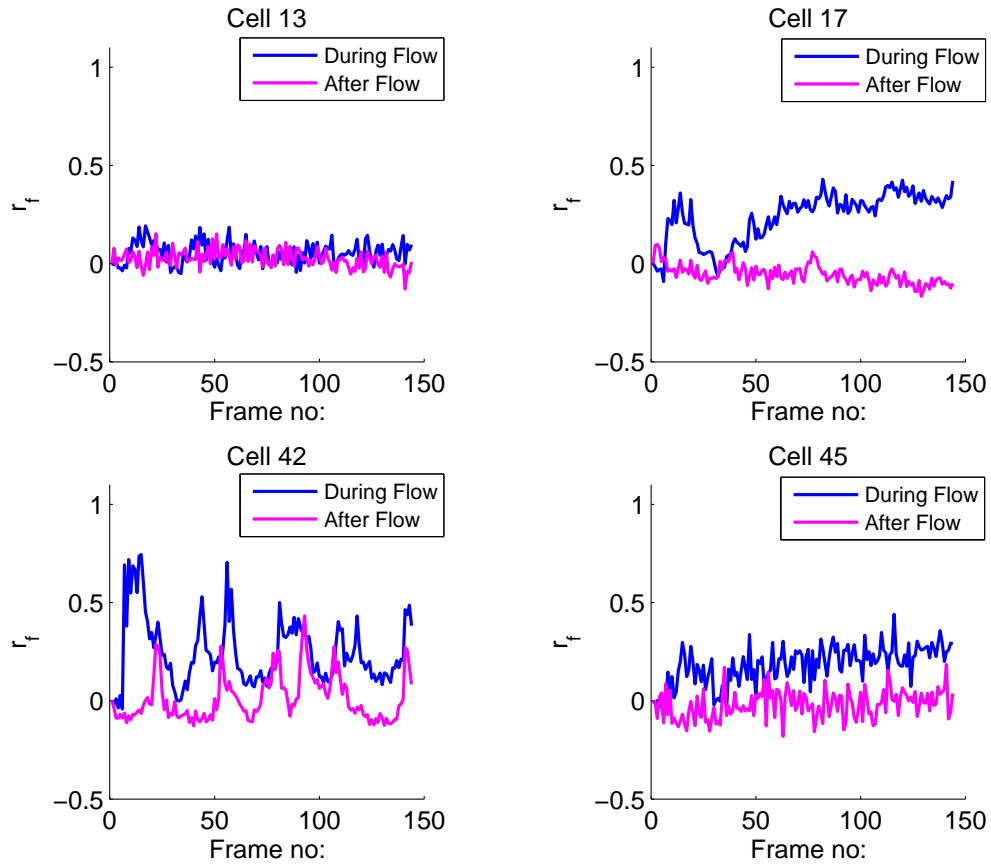
Figure 6.3.: Sample cell signals during flow and the signals after flow cessation, superimposed on top of each other.

6.3.1. Clustering for *In-vitro* Data

Clustering of the extracted signals was attempted to try to identify patterns in the calcium wave activities in both flow conditions. Cluster validation was done using the toolbox available in the file exchange version of *Matlab* [11]. In this toolbox, several validation indices were calculated for the hierarchical clustering algorithm. Validation indices for the signals "during flow" are given in Figures 6.4; whereas for the "after flow" condition, cluster validation results are presented in Figures 6.5.

The hierarchical clustering technique was applied to the signals obtained for all 50 cells studied in this dataset, as illustrated by the dendrograms in Figure 6.6 during flow and after flow cessation. Average signals under both flow conditions for all signals are given in Figures 6.6. Of a total of 50 cells studied in this dataset, cells formed three groups or clusters but some cells moved from one cluster to another; this suggests that some cells changed their Ca^{2+} concentration behaviour, while others remained the same under both conditions. The results of different cell groups during flow and after flow are presented in Figure 6.7, and their spatial locations are shown on the MIPT image in Figure 6.8.

The average signals of each cluster during flow do not change much but the change can be seen in after flow in average signals, but no significant patterns were observed through clustering technique. Therefore, instead of looking at the amplitudes of signals over time, it was then decided to look at the spatial patterns in cell transients, described in detailed in the next Section.

6.3.2. Calcium Transient Association in Cell Neighbourhood

Ca^{2+} transients were detected for 50 cells' "during flow" and "after flow" using an adaptation of the method described in [154]. In summary, before applying these algorithms, the signal was smoothed by using a low pass Sovitsky-Golay filter. After de-trending, the local maxima were determined as the times when the signal was larger than the signals for the two sampling times before and after. The signal was then zeroed by subtracting the intensity that corresponded to the peak of the histogram of local maxima, i.e., the mode of the intensity

122
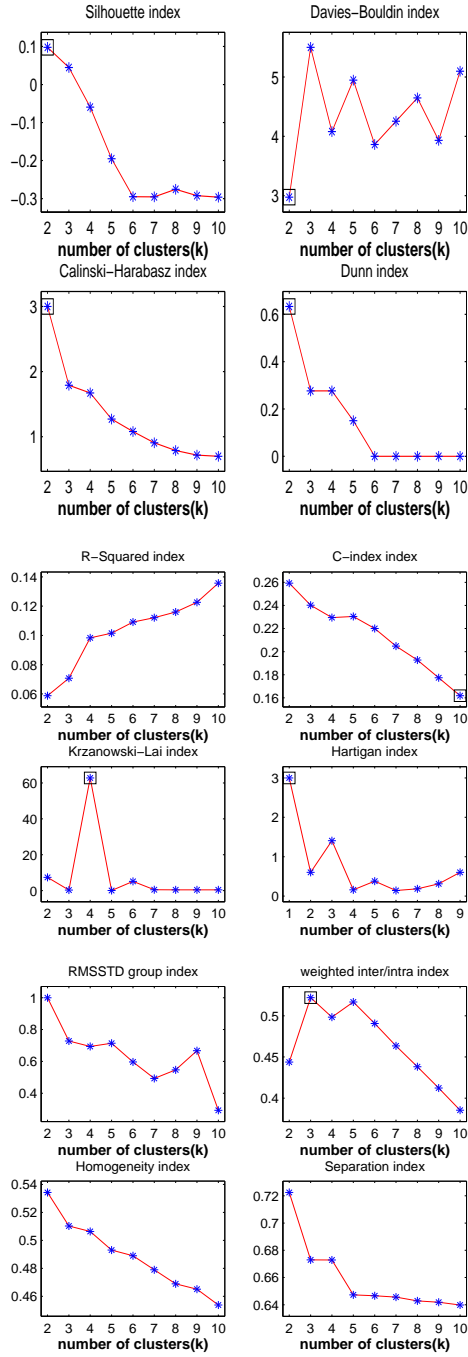
Figure 6.4.: Cluster validity indices found on signals during flow; an optimal number of clusters for each index is marked with a square box.
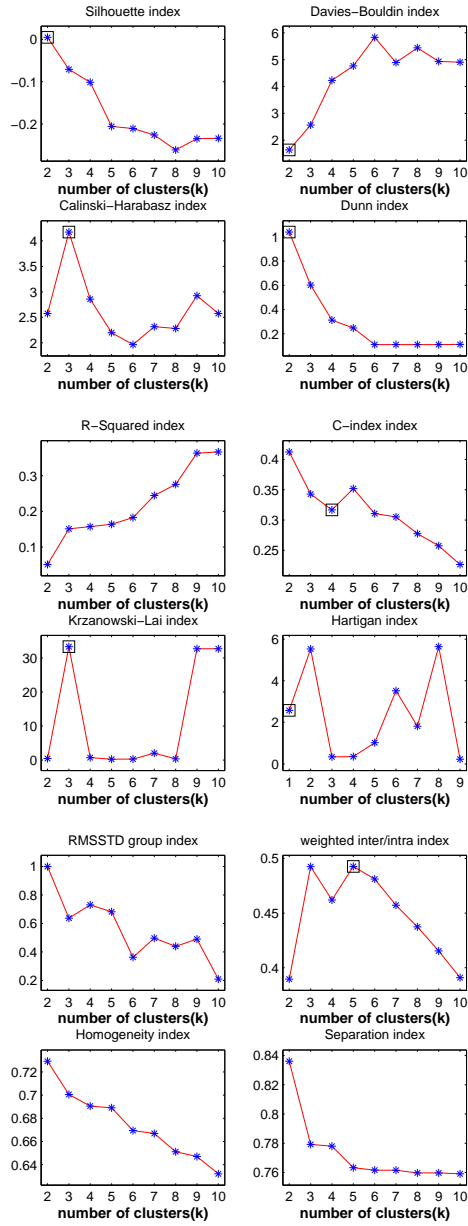
Figure 6.5.: Cluster validity indices found on signals after flow; an optimal number of clusters for each index is marked with a square box.
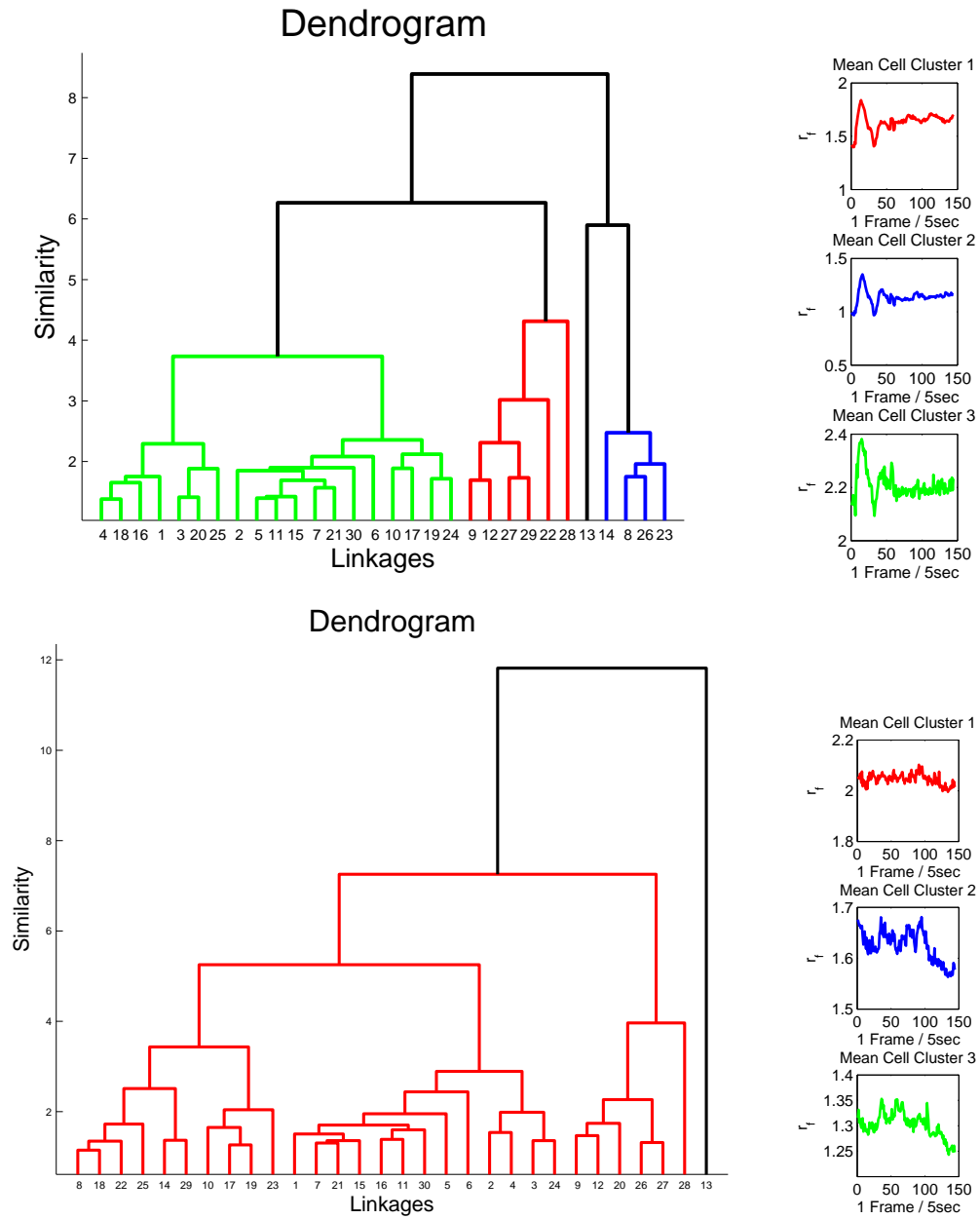
Figure 6.6.: Dendrograms and reconstructed "average" signals corresponding to the cluster centres for signal vectors during and after flow.
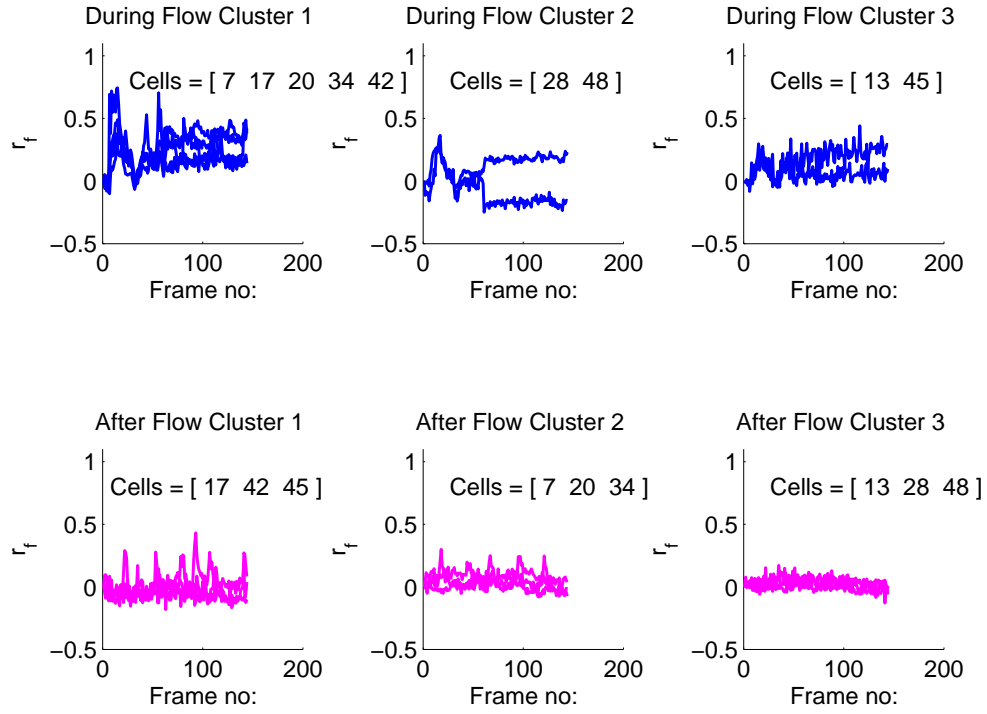
Figure 6.7.: Individual cell signals within each cluster, during flow and after flow. Several cells in this example [7,20, 28, 45, 48] changed from one shape cluster to another as a result of flow cessation.
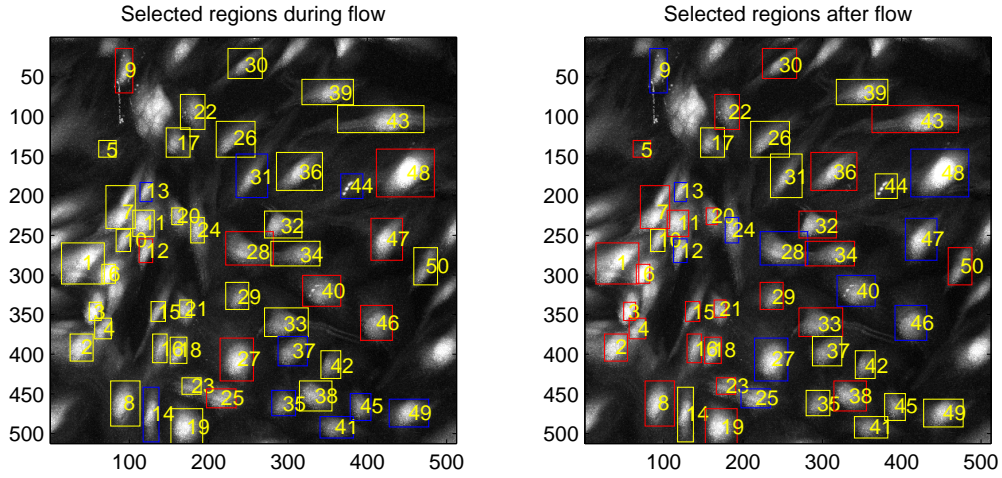
Figure 6.8.: Selected regions displaying different colours for the different clusters during flow and after flow.

maxima. The resulting data was then used as inputs to the transient detection algorithms, described below.

The transient detection algorithm first uses a Rosin thresholding algorithm, given in [146], and assumes that signal intensity values are composed of two populations (noise and transients in the signal), and produces two peaks (one main peak located at the lower end of the histogram relative to the secondary population). A line is drawn from this peak to the first empty bin of the histogram following the last occupied bin, and then a threshold between the two populations is set at the maximum perpendicular distance between the histogram and this line. Once it detects all the local maxima in the intensity signal, the Rosin algorithm is used to find a threshold intensity separating Ca^{2+} transients from lower amplitude maxima due to noise. The Rosin algorithm is again applied to the transients identified by the first step to find the threshold between the lower amplitude transients generated by a single Ca^{2+} transient and the higher amplitude transients generated by multiple transients that occurred within the duration of a single Ca^{2+} transient.

The second algorithm, described as the Template/Rosin method, uses a template of a Ca^{2+} transient by taking the ensemble average of the intensity signal

127

within a window of the detected transient times of the putative transients identified by the first stage of the Rosin method. Next, the correlation between the intensity signal and the template was calculated, and the peaks in the correlation were identified. The algorithm then works exactly in the same way as the previous algorithm but using the correlation signal rather than the intensity signal. The Rosin algorithm was used to separate the peaks due to transients from other small peaks due to noise. The transient peaks were then further separated using the Rosin algorithm into peaks resulting from single transients and peaks resulting from multiple transients. The amplitude of calcium transients was calculated as the magnitude of the intensity signal obtained from the Ca^{2+} fluorescence ratio signals at detected transient locations. Single and multiple fluorescence transients can be seen in the cells. A few samples of these are shown in Figure 6.9. Transient rate was calculated as the number of transients per cell over time during and after flow.

6.3.3. Transient Patterns in Cell Neighbourhood

Pixel-distance (using Euclidean distance) was calculated between each boundary pixel of the reference cell to all boundary pixels of each other cell in the database, as given in Algorithm 6.1 and the cell boundaries were overlaid (in red) for each cell, shown in Figure 6.10. If any pixel in any cell lies within 36 μm of any pixel in the reference cell, then that cell is identified as a "neighbouring cell" of the reference cell. This way, all the nearest neighbours were calculated for cells in the database. The total number of neighbouring cells varies for each cell.

To look for spatial correlations in Ca^{2+} transients amongst cell neighbourhoods, an algorithm for detecting Ca^{2+} transient patterns was developed and is given in Algorithm 6.2. Briefly, a reference cell was first selected. After occurrence of a transient from the reference cell, the nearest neighbours transient cells within a distance of 36 μm and within the current ("short time range cell-cell interaction") and two consecutive frames or 5-10 seconds at a rate of 5 frames per second ("long time range cell-cell interaction") were located. If two or more cell transients in this time range were detected amongst the neigh-
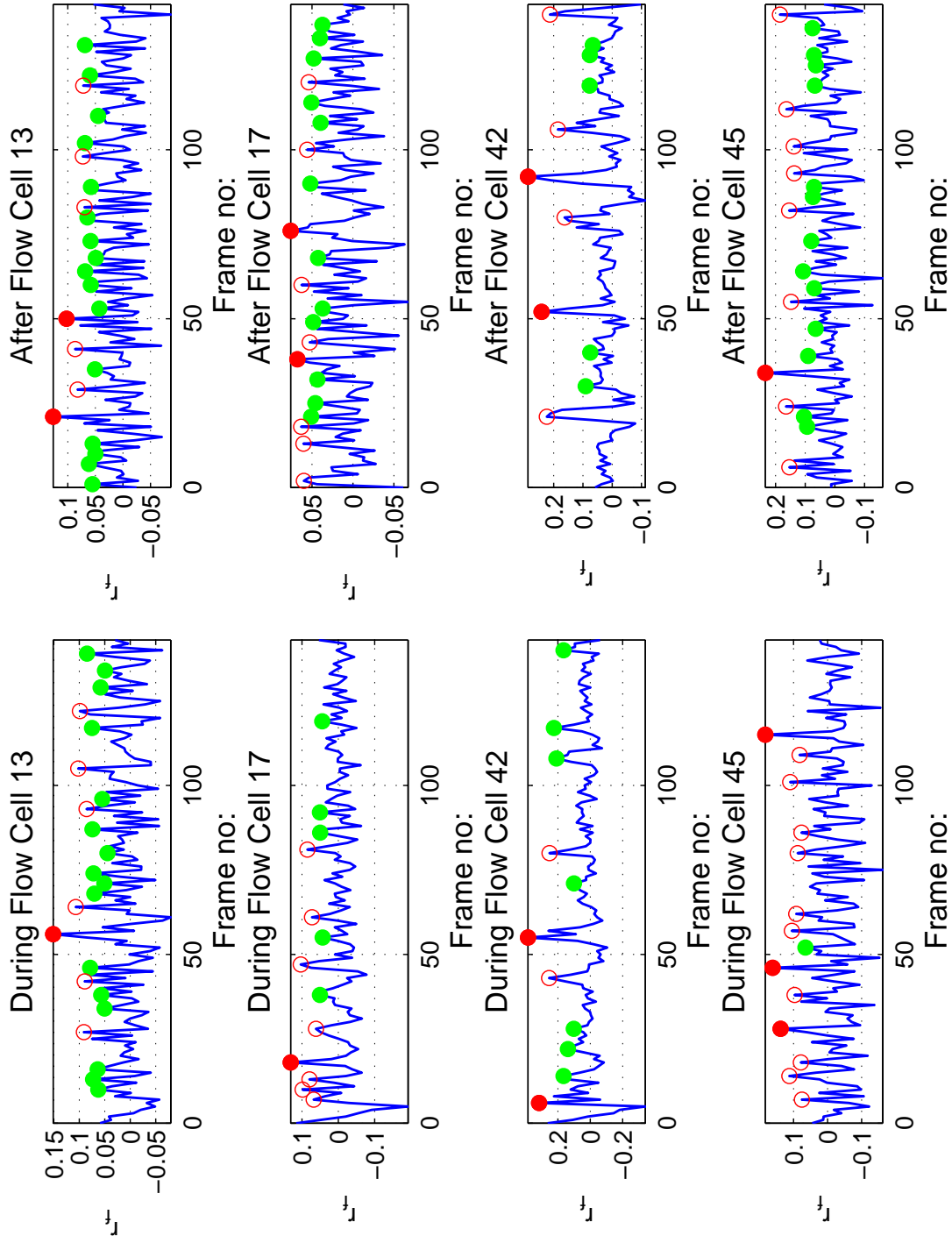
Figure 6.9.: Examples of cell Ca^{2+} signals during flow and after flow marked with red filled and unfilled are captured as transient signals and green dots are classified as noise in the signal.
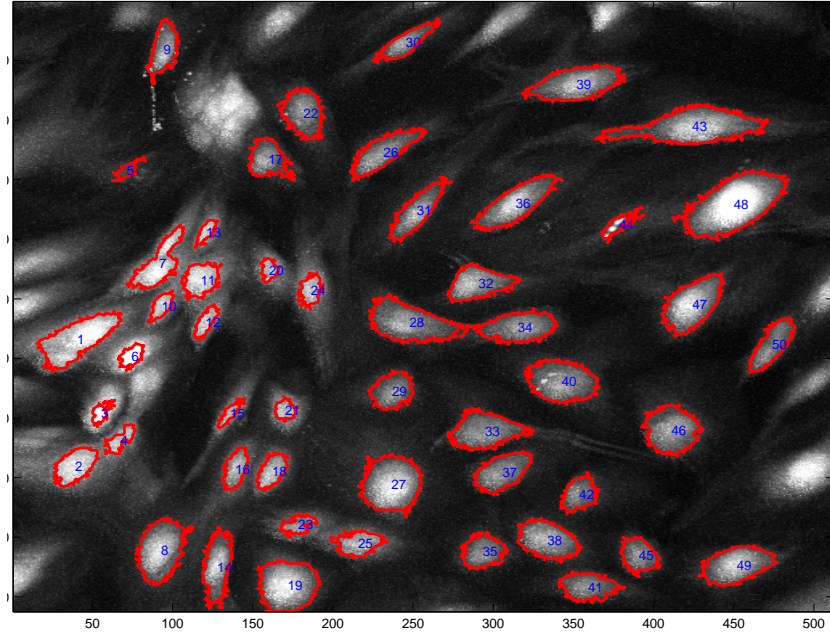
Figure 6.10.: Boundaries for each region displayed in (red) on maximum intensity projection over time image.

Algorithm 6.1 Algorithm to find cell neighbourhood

Initialize maximum(Max) distance for neighbour search;

i := 1: n; n := total number of regions;

while i $\leq$ n **do**

  Calculate all distances between each pixel of boundary of a region with each pixel of boundary of another region, yielding, Dist, a L1 $\times$ L2 distance matrix; L1 and L2 are the size (number of pixels) of boundaries of regions;

  MaxDist := any(Dist) $\geq$ Max;

  if MaxDist **then**

    N(i) := MaxDist;

  end if

end while

---

bours, then the nearest neighbour amongst the transient cells was selected as a new potential reference cell for a further search. In this way, the process continued until all frames were analysed, and the transient pattern was found for the reference cell; the same process was then repeated for all the cells.

Algorithm 6.2 Algorithm for identifying calcium transient associations in cell neighbourhood

---

j := 1: t; t := total number of frames;
i := 1: n; n := total number of regions;
reference cell := i;
TRANSIENTS := All cells transient timings;
Transient := TRANSIENTS(i);
while j ≤ t **do**
  if any(TRANSIENTS) == j **then**
    find(TRANSIENTS(NN)); NN := nearest neighbours of ith region;
    if any(TRANSIENTS(NN)) == j **then**
      Display frame j connection in red;
      reference cell := cell which transient at j;
    end if
    if any(TRANSIENTS(NN)) == j+1 **then**
      Display frame j+1 connection in blue;
      reference cell := cell which transient at j+1;
    end if
    if any(TRANSIENTS(NN)) == j+2 **then**
      Display frame j+2 connection in green;
      reference cell := cell which transient at j+2;
    end if
  end if
end while

Different patterns in calcium fluorescence obtained during and after flow are depicted in Figure 6.3. It is clear that a large heterogeneity in the response to flow exists in endothelial cells in culture, as there are endothelial cells where Ca^{2+} concentrations remained unchanged, while in other cells the amplitude of the Ca^{2+} transients increased Figure 6.11. The relative amplitude of the Ca^{2+} transients did not change significantly after flow cessation (1.76 ± 0.34 during flow and 1.72 ± 0.36 after flow cessation).
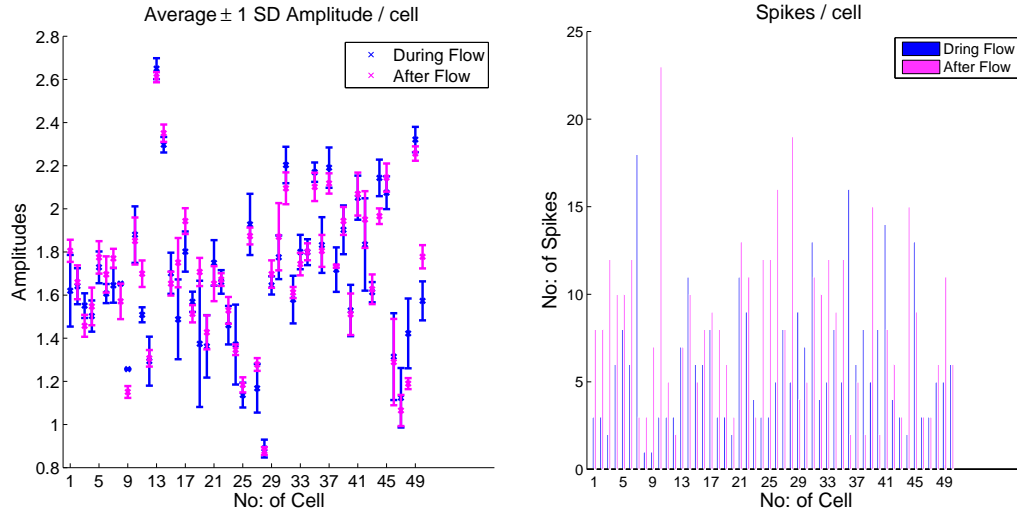
Figure 6.11.: During flow and after flow amplitude distribution of cells and transient distribution of cells.

The transient amplitude and rate distributions for each cell are also given in Figure 6.11, where it can be noted that 84 % of all cells were transient at a higher rate after flow cessation. If we assume that the transient rate distributions before and after flow may be modelled by separate Poisson distributions, then the maximum likelihood estimate of the Poisson parameters changed from 6.06 during flow to 8.24 after flow cessation confirming the findings from Figure 6.11.

Using the transient searching method algorithm 6.2, we found that the transient endothelial cells formed spatial and temporal patterns with their neighbouring cells. Such patterns are represented in Figures 6.12 and 6.13 at different timing points during flow and after flow cessation. It is clear that most cells interact with their direct neighbours only ("short range interaction") while some cells appear to have ("long range interaction"), and cells form circular pattern after flow is ceased.
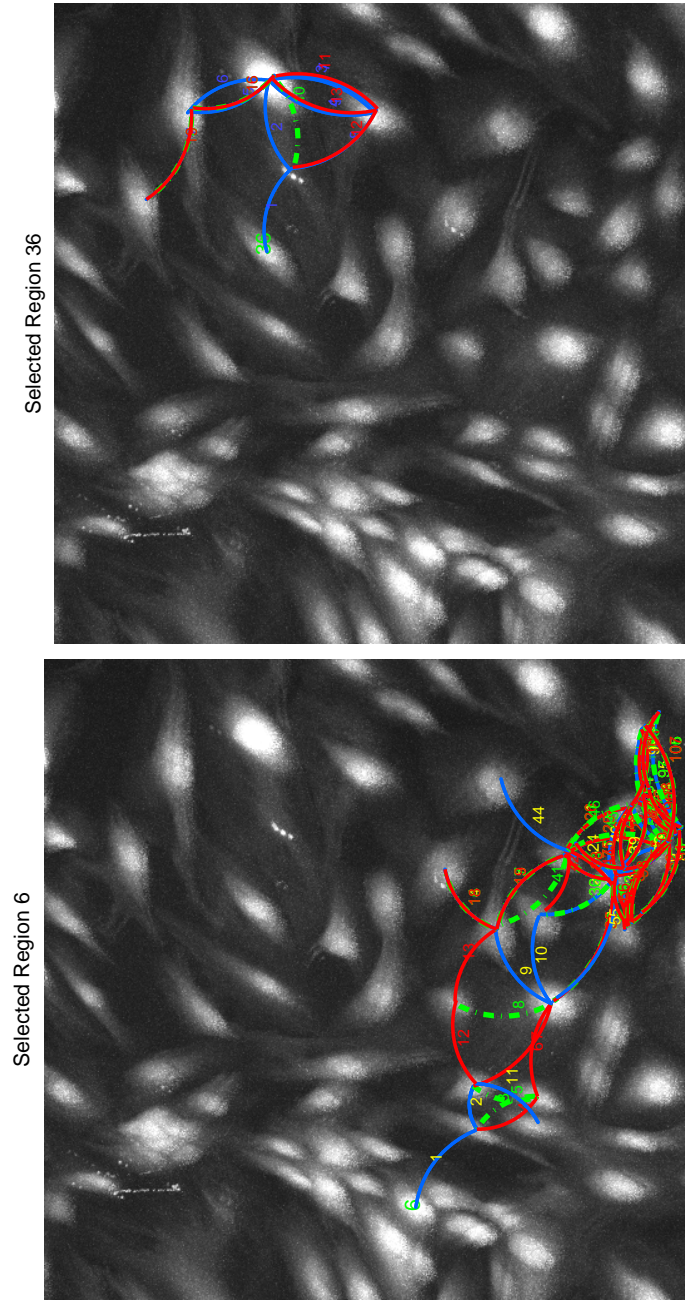
Figure 6.12.: Cell transients in relation to neighbourhood. Here, cells6 and 36 during flow are shown. Numbers indicate the centroid locations of each Transient cell. Red lines represent transient in neighbour cells within the same frame as the reference cell; while green and blue lines show neighbour cells Transient, respectively 1 and 2 frames later than the reference cell.
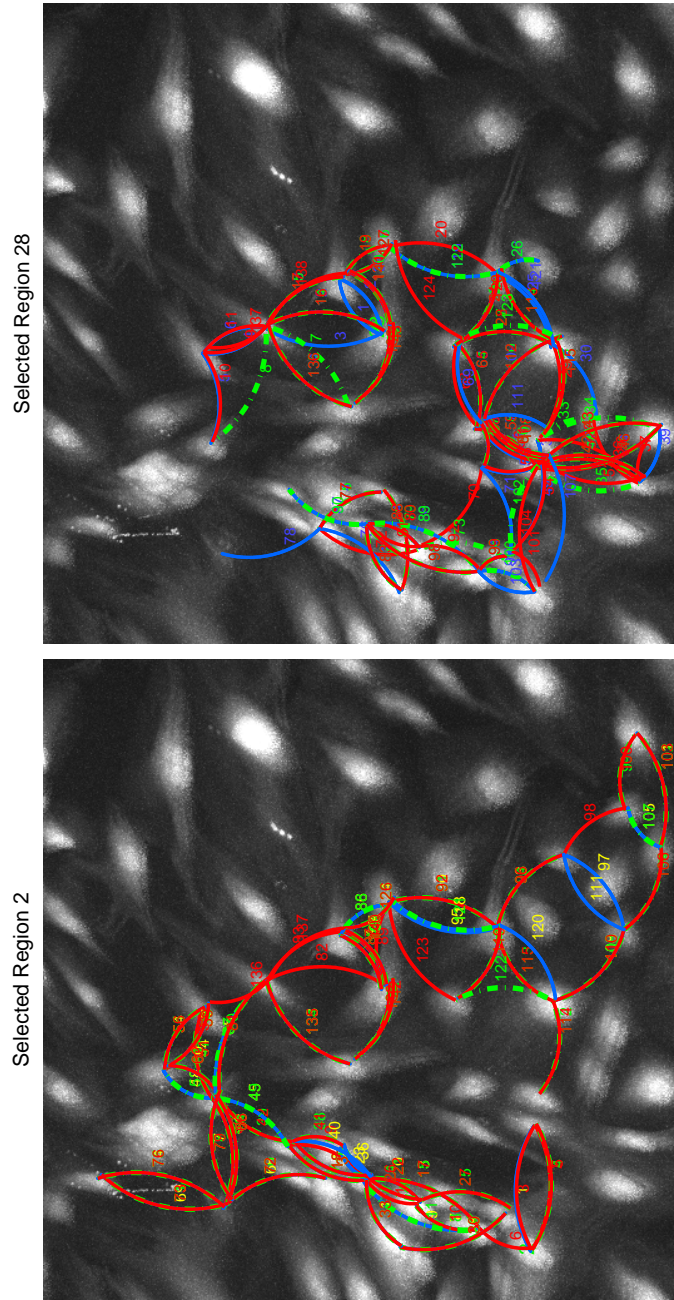
Figure 6.13.: Cell transients in relation to neighbourhood. Here, cells 2 and 28 after flow are shown. Numbers indicate the centroid locations of each Transient cell. Red lines represent transient in neighbour cells within the same frame as the reference cell; while green and blue lines show neighbour cells Transient, respectively 1 and 2 frames later than the reference cell.

6.4. Discussion

This chapter ECs behaviour was investigated in a controlled flow environment and different techniques were developed in this thesis and applied to detect calcium transients in cells and to analyse the patterns of these transients between neighbouring cells.

A new method is presented for detecting the effect of flow on the spatial and temporal patterns of calcium activity among neighbouring cells. On average, the probability of a cell firing in a given frame is 0.05. However, a small number of cells had a much higher probability (≥ 0.25) of firing if a neighbouring cell fired during the current frame period. After flow cessation, cells which had firing probabilities higher than 0.05, formed intricate spatial patterns among neighbouring cells.

Several groups have postulated that ATP transport may be involved in the process of flow sensing by the endothelial cell [46, 49, 130]. Interestingly ATP is released from endothelial cells under shear stress [23] but is - at the same time - degraded at the cell surface by ATPase. Hence, some investigators suggested the modulation of ATP concentration at the cell surface by flow to explain the shear dependent calcium response. Does a mass transport process, in particular of ATP, really account for the flow dependencies of calcium transient observed? Controversial results can be found in the literature. The original experiment of Ando [85] revealed that ATP is necessary for calcium transients but in other experiments in which ATP release was blocked or ATP effectively degraded, calcium transients could still be observed [152]. The propagation of the calcium wave to neighbouring cells can be significantly suppressed by a purinergic receptor blocker, implying that ATP participates in the propagation process [75]. Heterogeneity in cell responses could be due to a non-uniform distribution of purinoreceptors. ATP is able to activate the nearby endothelial cells or can be transported by blood flow downstream. Hong and others suggested a mechanism by which ATP acts as an autocrine mediator, integrating individual cell responses and coordinating vascular function [50, 75]. Further experiments investigating these calcium waves suggested they are transmitted by both ATP via purinoreceptors and IP3 via gap junction (connexin 40) [166].

135

6. Experiments on In-vitro Cells (Temporal Data)

The current experiments add further to the above mentioned controversies. If mass transport of ATP plays a role in calcium transients, then flow would affect propagation of calcium patterns in such a way that they would follow the direction of the streamlines of the flow field. We did find that the presence or absence of flow affected the calcium transient patterns, but rather unexpectedly the calcium transient patterns were very complex, with patterns, perpendicular to stream lines and even retrograde to the stream lines. These observations therefore suggest that the effect of flow on calcium transients and their cell-cell propagation is more likely due to calcium propagation through intercellular gap and tight junctions and to a smaller amount over stream lines. This then suggests that transport of species along streamlines is of less importance, at least in the current experiments.

In conclusion, cell transient appears synchronised in multi-cellular groups, forming complex patterns that bear little relationship with streamlines of the velocity field, but that do change with shear stress. The patterns suggest a new functional mechanism in flow-mediation of cell-cell calcium signalling.

7. Conclusions and Future Work

7.1. Summary

At the beginning of this thesis, the biological motivation behind the study of the endothelial cell and of heterogeneity in its functions, structure and its spatial patterns is discussed.

This research is aimed at developing methods for the automated analysis of spatial patterns of endothelial cells (ECs), and the functional behaviour of endothelial cells under different flow conditions and in different species. This would enable researchers to analyse large amounts of microscopy data in order to understand the patterns of formation of atherosclerotic lesions.

For many types of automatic analysis, the first step is to segment the data, and then group them into different classes. Various related segmentation and classification techniques are briefly reviewed in Chapter 2 in order to establish a context to present the different algorithms developed in this thesis.

The first step is the detection of the cell nucleus and cell boundaries. In Chapter 3 different algorithms for segmentation are presented for endothelial nuclei (ECn) and, particularly for endothelial cell (ECs) boundaries.

Chapter 4 analysed changes in patterns for stress affected morphological parameters of endothelial cells (ECs) and endothelial nuclei (ECn) and how they varied with age in rabbit data.

In Chapter 5, a hypothesis is suggested on ECs and ECn shapes. This could be used to look at the other cell shape measurements instead of just looking at length-to-width ratio and angle of orientation and their relationship to the functions of the cells. To test this hypothesis, two shape analysis techniques are developed. This may help in future to understanding and analysis of the morphological variations in the shapes of endothelial cells and endothelial nu-

137

clei.

Chapter 6 investigated the behaviour of ECs in a controlled flow environment and applied techniques developed in this thesis to detect heterogeneity among cells, particularly the formation of intricate patterns in single-cell transients using Ca^{2+} sensitive florescence probes analysis.

7.2. Research Outcomes

The research outcomes and new methods explored in the thesis may be categorised into the following headings:

1. Candidate shape model

2. Patch-based feature vector for cell segmentation

3. Morphological parameter based association between *in-vivo* endothelial cells and endothelial nuclei

4. Age-related changes in patterns of *in-vivo* endothelial cell data

5. A hypothesis on *in-vivo* endothelial cell data

6. A new algorithm for Ca^{2+} transient pattern detection

These findings are discussed in detail below:

7.2.1. Candidate Cell Shape Model

An approach is introduced in Chapter 3 to fit an equivalent ellipse to the cell boundary using biologically inspired wavelet models. This approach fits a parametric model to describe each cell boundary by matching a phase-invariant orientation field of a candidate cell shape with that acquired from the image. This approach succeeded on good quality images, but failed on images with weak cell boundaries.

7.2.2. Patch-based Feature Vector for Endothelial Cell Segmentation

The first objective was the detection of the cell nucleus and its boundary. The detection of the nucleus was an easier task compared to the detection of cell boundaries due to low-contrast and low signal-to noise ratio of the cell boundaries. To perform the segmentation of cell boundaries, a new patch-based 33-dimensional feature vector for cell segmentation is introduced in Chapter 3, since for the kernel-based support vector machines an appropriate feature set is vital. Phase invariant orientation patches also turn out to contribute strongly to the feature vector, and results presented suggest that there is no penalty for using a large feature vector. A caution considering feature selection is that certain features should be grouped together: patch-based data are a case in point - forward feature selection tends not to group data in this way; however, it is the combination of pixel information in the neighbourhood of a candidate pixel that is important. Whilst this poses a problem of generalisation from limited training samples with traditional classifiers, the clear implication from the results shown here is that the complexity of the resulting decision boundary does not present an obstacle to good generalisation for an support vector machine with a Gaussian kernel. A key feature of this approach is its extendibility to other, similar segmentation problems in different imaging modalities.

The method has some limitations; it does not perform well on image patches with very large brightness variation or where cell boundaries are not visible by eye. However, that scenario would pose significant problems for other segmentation approaches as well. It also requires a manual labelling of pixels of one example as input in the training process. This necessitates selecting a patch with a representative variation in intensity to yield a satisfactory generalized model.

Another aspect of this algorithm which could be improved is the selection of kernels. In this thesis, training has been done using only kernels already available in a standard SVM package; performance might be improved by developing a new kernel for such data, which is beyond the scope this research. Nevertheless, the method enabled us to explore the relationships between blood

flow and cell behaviour, between cell nuclear shape and cell boundary shape, and between aspect ratio and cell orientation in different regions around arterial branches, by facilitating the analysis of large numbers of cells.

7.2.3. Morphological Parameter Based Association between *In-vivo* Cells and Nuclei

A new algorithm is introduced in Chapter 4 to find the association between cells and their respective nuclei using various morphological measures and the nuclei centroids as *a priori* information. The algorithm produces a complete set of associated cells and their respective nuclei for further study of the relationship between shear-stress dependent morphological measures.

7.2.4. Age-related Changes in Patterns

In this thesis, additional evidence for similar behaviour of endothelial cells and their nuclei is obtained *in vivo*. Nuclear and cell orientations around branches in immature and mature rabbits are broadly similar and found to be highly correlated, suggesting that the pattern of flow direction near the wall stayed approximately constant with age. Importantly, the correlation is positive rather than negative: high lesion frequency correlates with greater nuclear elongation, consistent with lipid deposition occurring in regions experiencing high wall shear stress at both ages. The relevance to human lesions, which also switch from a downstream to a lateral or upstream location around the intercostal branch ostia with age [180], remains to be determined [25].

In cell and nuclear length-to-width ratios, the most striking features are low values in immature aortas, with high values in mature aortas. However, the relation is less strong for length-to-width ratios than for alignment, so other influences on elongation, such as pulsatility or wall stiffness (which varies around branches and changes with age), although not considered herein, cannot be entirely dismissed [25].

7.2.5. A Hypothesis on *In-vivo* Endothelial Cell Data

A hypothesis on cell data is presented to look at the data with different dimensions; rather than just looking at simple morphological parameters. In order to explore this hypothesis, in this thesis, two new shape analysis techniques shape-descriptor analysis, and a shape-boundary contour analysis, are presented for cell and nuclear data. These techniques are developed to detect shape variations in the data given in Chapter 5. Initial analysis on nuclear data clearly show variation in shapes of immature and mature animal nuclei data. The representation of these shapes with respect to the atherosclerotic lesion sites and changes in patterns with age are yet to be determined. Future work will be to study the significance of this finding.

7.2.6. Calcium Transient Patterns in *In-vitro* Data

A new method is presented for detecting the effect of flow on the spatial and temporal patterns of calcium activity among neighbouring cells. On average, the probability of a cell firing in a given frame is 0.05. However, a small number of cells had a much higher probability (≥ 0.25) of firing if a neighbouring cell fired during the current frame period. After flow cessation, cells which had firing probabilities higher than 0.05 formed intricate spatial patterns among neighbouring cells.

Transient rate changed with flow, but more importantly, local patterns are discernible, which changed after stopping the flow. The patterns are not compatible with ATP release and suggest a new, unknown mechanism must be involved. In conclusion, cell transients appear to be synchronised in multi-cellular structures forming spatial and temporal patterns that change with flow.

7.3. Suggestion for Future Work

One future plan is to develop a machine learning algorithm and its code base for automatic generation of application-specific features and development of the kernel.

As the relation was less strong for length-to-width ratios than for alignment,

so other mechanical forces influence on elongation may be involved. There is a need to study and identify which mechanical forces are acting on cells and nuclei in order to explain this smaller correlation between cell and nuclei in length-to-width ratios. There seems to be some patterns in nuclear alignment in whole aortas, as shown in Figure 1 in [25]; it can also be an area of future work to find out how and why these patterns have formed.

To confirm the cell shape results, a huge number of cells representing many animals will be needed, and considerable work is required to explore both the statistical and biological significance of the results.

In the future, single cell probe analysis will lead to a better understanding of gene expression patterns at cellular-level and relating these genetic properties to cell shape and calcium transients will be of great interest.

Bibliography

[1] http://neuro.debian.net/datasets.html.

[2] https://www.rad.upenn.edu/sbia/.

[3] http://www.bd.com/.

[4] http://www.bhf.org.uk/.

[5] http://www.biophysics.com/ibidi/instruments/ibidiinstruments.html.

[6] http://www.cma.mgh.harvard.edu/ibsr/.

[7] http://www.dojindo.com/home/.

[8] http://www.dtreg.com/pnn.htm.

[9] http://www.mathworks.com/matlabcentral/fileexchange/22423-ellipse-fit.

[10] The American Heritage Science Dictionary. Houghton Mifflin Harcourt Publishing Company, 2010.

[11] J. Abonyi. Clustering toolbox. http://www.mathworks.de/matlabcentral/fileexchange/7486, 2005.

[12] P. Adiga, B. Chaudhuri, and K. Rodenacker. Semi-automatic segmentation of tissue cells from confocal microscope images. *Proceedings of the 13th International Conference on Pattern Recognition*, 3:494–497, 1996.

[13] R. K. Aggarwal and J. W. Bacus. A multi-spectral approach for scene analysis of cervical cytology smears. *Journal of Histochemistry and Cytochemistry*, 25:668–80, 1977.

Bibliography

[14] W. C. Aird. Endothelial cell heterogeneity. *Critical Care Medicine*, 31:221–230, 2003.

[15] W. C. Aird. Endothelial cell heterogeneity and atherosclerosis. *Current Atherosclerosis Reports*, 8:69–75, 2006.

[16] S. L. Al-Musawi, J. Bishton, J. Dean, S. Williams, S. G. Cremers, and P. D. Weinberg. Evidence for a reversal with age in the pattern of near-wall blood flow around aortic branches. *Atherosclerosis*, 172:79–84, 2004.

[17] E. Backer and A. K. Jain. A clustering performance measure based on fuzzy set decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3:66 –75, 1981.

[18] H. S. Badeer. Hemodynamics For Medical Students. *Advances in Physiology Education*, 25:44–52, 2001.

[19] J. Bezdek and N. Pal. Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 28:301–315, 1998.

[20] J. C. Bezdek, L. O. Hall, and L. P. Clarke. Review of mr image segmentation techniques using pattern recognition. *Medical Physics*, 20:1033–1048, 1993.

[21] A. A. Bharath and J. Ng. A steerable complex wavelet construction and its application to image denoising. *IEEE Transactions on Image Processing*, 14:948–959, 2005.

[22] W. Bogner, S. Gruber, K. Pinker, G. Grabner, A. Stadlbauer, M. Weber, E. Moser, T. H. Helbich, and S. Trattnig. Diffusion-weighted MR for Differentiation of Breast Lesions at 3.0 T: How Does Selection of Diffusion Protocols Affect Diagnosis? *Radiology*, 253:341–351, 2009.

[23] S. Boitano, E. Dirksen, and M. Sanderson. Intercellular propagation of calcium waves mediated by inositol trisphosphate. *Science*, 258:292–295, 1992.

Bibliography

[24] A. R. Bond. *Effect of age and species on blood flow patterns at arterial branches in relation to atherosclerosis.* PhD thesis, University of London, Imperial College London, United Kingdom, 2007.

[25] A. R. Bond, S. Iftikhar, A. A. Bharath, and P. D. Weinberg. Morphological evidence for a change in the pattern of aortic wall shear stress with age. *Arterioscler Thromb Vasc Biol*, 31:543–50, 2011.

[26] A. R. Bond, C.-W. Ni, H. Jo, and P. D. Weinberg. Intimal cushions and endothelial nuclear elongation around mouse aortic branches and their spatial correspondence with patterns of lipid deposition. *American Journal of Physiology - Heart and Circulatory Physiology*, 298:536–544, 2010.

[27] R. L. Cahn, R. S. Poulsen, and G. Toussaint. Segmentation of cervical cell images. *Journal of Histochemistry & Cytochemistry*, 25:681–8, 1977.

[28] P. D. Caie, R. E. Walls, A. Ingleston-Orme, S. Daya, T. Houslay, R. Eagle, M. E. Roberts, and N. O. Carragher. High-Content Phenotypic Profiling of Drug Response Signatures across Distinct Cancer Cells. *Molecular Cancer Therapeutics*, 9:1913–1926, 2010.

[29] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[30] S. D. Cataldo, E. Ficarra, A. Acquaviva, and E. Macii. Automated segmentation of tissue images for computerized IHC analysis. *Computer Methods and Programs in Biomedicine*, 100:1–15, 2010.

[31] V. Chalana, D. Linker, D. Haynor, and Y. Kim. A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Transactions on Medical Imaging*, 15:290–298, 1996.

[32] C.-C. Chang and C.-J. Lin. Libsvm: a library for Support Vector Machines. *Software available at* http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[33] C. Chen, H. Lu, and Y. Huang. Cell-based dual snake model: a new approach to extracting highly winding boundaries in the ultrasound images. *Ultrasound in Medicine and Biology*, 28:1061–1073, 2002.

[34] J. Cheng and S.-W. Foo. Boundary detection in echocardiographic images using markovian level set method. *IEICE Trans. Inf. Syst.*, 90:1292–1300, 2007.

[35] S. Chien. Mechanotransduction and endothelial cell homeostasis: the wisdom of the cell. *Am J Physiol Heart Circ Physiol*, 292:209–224, 2007.

[36] M. Clark. Quantitative shape analysis: A review. *Mathematical Geology*, 13:303–320, 1981.

[37] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437:699–706, 2005.

[38] J. F. Cornhill, M. J. Levesque, E. E. Herderick, R. M. Nerem, J. W. Kilman, and J. S. Vasko. Quantitative study of the rabbit aortic endothelium using vascular casts. *Atherosclerosis*, 35:321–337, 1980.

[39] C. Cortes. Invited talk: Can learning kernels help performance? *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009.

[40] L. Costa and R. Cesar. *Shape Analysis and Classification: Theory and Practice*. CRC Press, Boca Raton, Florida, 2001.

[41] M. Costa, H. Delingette, S. Novellas, and N. Ayache. Automatic segmentation of bladder and prostate using coupled 3D deformable models. *Medical Image Computing and Computer-Assisted Intervention*, 4791:252–260, 2007.

[42] N. Cristianini. http://www.support-vector.net/erice.pdf.

[43] L. da F. Costa, S. F. dos Reis, R. A. T. Arantes, A. C. R. Alves, and G. Mutinari. Biological shape analysis by digital curvature. *Pattern Recognition*, 37:515–524, 2004.

[44] L. da F. Costa and T. J. Velte. Automatic characterization and classification of ganglion cells from the salamander retina. *The Journal of Comparative Neurology*, 404:33–51, 1999.

[45] S. R. Daniels, F. R. Greer, and the Committee on Nutrition. Lipid screening and cardiovascular health in childhood. *Pediatrics*, 122:198–208, 2008.

[46] P. F. Davies. Flow-mediated endothelial mechanotransduction. *Physiol. Rev.*, 75:519–560, 1995.

[47] P. F. Davies. Hemodynamic shear stress and the endothelium in cardiovascular pathophysiology. *Nature clinical practice. Cardiovascular medicine*, 6:16–26, 2009.

[48] M. B. Dillencourt, H. Samet, and M. Tamminen. A general approach to connected-component labeling for arbitrary image representations. *J. ACM*, 39:253–280, 1992.

[49] R. O. Dull and P. F. Davies. Flow modulation of agonist (ATP)-response (Ca2+) coupling in vascular endothelial cells. *Am J Physiol Heart Circ Physiol*, 261:149–154, 1991.

[50] T. Duza and I. H. Sarelius. Localized transient increases in endothelial cell Ca2+ in arterioles in situ: implications for coordination of vascular function. *Am J Physiol Heart Circ Physiol*, 286:2322–2331, 2004.

[51] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*, 21:1552–1563, 2002.

[52] Y. Fan, D. Shen, and C. Davatzikos. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM.

Medical Image Computing and Computer-Assisted Intervention, 3749:1–8, 2005.

[53] J. T. Flaherty, J. E. Pierce, V. J. Ferrans, D. J. Patel, W. K. Tucker, and D. L. Fry. Endothelial Nuclear Patterns in the Canine Arterial Tree with Particular Reference to Hemodynamic Events. *Circ Res*, 30:23–33, 1972.

[54] Y.-L. Fok, J. Chan, and R. Chin. Automated analysis of nerve-cell images using active contour models. *IEEE Transactions on Medical Imaging*, 15:353–368, 1996.

[55] D. Foran, D. Comaniciu, P. Meer, and L. Goodell. Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *Information Technology in Biomedicine, IEEE Transactions on*, 4:265–273, 2000.

[56] C. Fraley and A. E. Raftery. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, 41:578–588, 1998.

[57] H. Freeman and L. Davis. A corner-finding algorithm for chain-coded curves. *IEEE Transactions on Computers*, 26:297–303, 1977.

[58] J. Freixenet, X. Muoz, D. Raba, J. Mart, and X. Cuf. Yet another survey on image segmentation: Region and boundary information integration. 2352:21–25, 2002.

[59] K. Fu and J. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.

[60] C. Garlanda and E. Dejana. Heterogeneity of endothelial cells : Specific markers. *Arterioscler Thromb Vasc Biol*, 17:1193–1202, 1997.

[61] A. R. S. Garraffoni and M. de Garcia Camargo. First application of morphometrics in a study of variations in uncinial shape present within the terebellidae (polychaeta). *Zoological Studies*, 45:75–80, 2006.

[62] Y. Gavet and J.-C. Pinoli. Visual perception based automatic recognition of cell mosaics in human corneal endothelium microscopy images. *Image Anal Stereol*, 27:53–61, 2008.

[63] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P. Arapantoni Dadioti, and G. Nikiforidis. An image-analysis system based on Support Vector Machines for automatic grade diagnosis of brain-tumour astrocytomas in clinical routine. *Informatics for Health and Social Care*, 30:179–193, 2005.

[64] K. Goshima, A. Masuda, and K. Owaribe. Insulin-induced formation of ruffling membranes of KB cells and its correlation with enhancement of amino acid transport. *The Journal of Cell Biology*, 98:801–809, 1984.

[65] G. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. 1995.

[66] L. Gupta, M. R. Sayeh, and R. Tammana. A neural network approach to robust shape classification. *Pattern Recognition*, 23:563–568, 1990.

[67] L. Hall, A. Bensaid, L. Clarke, R. Velthuizen, M. Silbiger, and J. Bezdek. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *Neural Networks, IEEE Transactions on*, 3:672–682, 1992.

[68] G. K. Hansson. Inflammation, atherosclerosis, and coronary artery disease. *New England Journal of Medicine*, 352:1685–1695, 2005.

[69] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–132, 1985.

[70] N. Harder, B. Neumann, M. Held, U. Liebel, H. Erfle, J. Ellenberg, R. Eils, and K. Rohr. Automated recognition of mitotic patterns in fluorescence microscopy images of human cells. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1:1016–1019, 2006.

Bibliography

[71] G. Helmlinger, B. C. Berk, and R. M. Nerem. Calcium responses of endothelial cell monolayers subjected to pulsatile and steady laminar flow differ. *Am J Physiol Cell Physiol*, 269:367–375, 1995.

[72] G. Helmlinger, B. C. Berk, and R. M. Nerem. Pulsatile And Steady Flow-Induced Calcium Oscillations In Single Cultured Endothelial Cells. *Journal Of Vascular Research*, 33:360–369, 1996.

[73] G. Helmlinger, R. V. Geiger, S. Schreck, and R. M. Nerem. Effects of pulsatile flow on cultured vascular endothelial cell morphology. *Journal of Biomechanical Engineering*, 113:123–131, 1991.

[74] G. Holzapfel, T. Gasser, and R. Ogden. A new constitutive framework for arterial wall mechanics and a comparative study of material models. *Journal of Elasticity*, 61:1–48, 2000.

[75] D. Hong, K. Barbee, D. Buerk, and D. Jaron. Heterogeneous cytoplasmic calcium response in microvascular endothelial cells. *Annual International Conference of the Engineering in Medicine and Biology Society*, 1:7493–7496, 2005.

[76] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19:203–210, 2000.

[77] S. L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *J. ACM*, 23:368–388, 1976.

[78] M.-K. Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8:179–187, 1962.

[79] S. Hua and Z. Sun. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, 17:721–728, 2001.

[80] K. Huang and R. Murphy. Automated classification of subcellular patterns in multicell images without segmentation into single cells. *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, 1:1139–1142, 2004.

Bibliography

[81] K. Huang, M. Velliste, and R. F. Murphy. Feature reduction for improved recognition of subcellular location patterns in fluorescence microscope images. *Manipulation and Analysis of Biomolecules, Cells, and Tissues*, 4962:307–318, 2003.

[82] I. Hunter, J. Soraghan, and T. McDonagh. Fully automatic left ventricular boundary extraction in echocardiographic images. *Computers in Cardiology*, 1:741–744, 1995.

[83] S. Iftikhar and A. A. Bharath. Cell boundary analysis using radial search for dual staining techniques. *Proc. SPIE Medical Imaging*, 7259, 2009.

[84] G. M. Ilias and E. P. Zafiropoulos. Characterization of digital medical images utilizing Support Vector Machines. *BMC Medical Informatics and Decision Making*, 4:4, 2004.

[85] M. Isshiki, J. Ando, R. Korenaga, H. Kogo, T. Fujimoto, T. Fujita, and A. Kamiya. Endothelial Ca2+ waves preferentially originate at specific loci in caveolin-rich cell edges. *Proceedings of the National Academy of Sciences of the United States of America*, 95:5009–5014, 1998.

[86] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31:264–323, 1999.

[87] A. Jalba, M. Wilkinson, J. Roerdink, M. Bayer, and S. Juggins. Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Machine Vision and Applications*, 16:217–228, 2005.

[88] M. Jean. An unsupervised multiscale approach for the dynamic contour-based boundary detection issue in ultrasound imagery, 2008.

[89] S. J. Jelonek J. Feature subset selection for classification of histological images. *Artificial Intelligence in Medicine*, 9:227–239, 1997.

[90] I. Kalatzis, D. Pappas, N. Piliouras, and D. Cavouras. Support Vector Machines based analysis of brain spect images for determining cerebral

151

abnormalities in asymptomatic diabetic patients. *Informatics for Health and Social Care*, 28:221–230, 2003.

[91] L. F. Kaspersen J.H., Lango T. Wavelet-based edge detection in ultrasound images. *Ultrasound in Medicine and Biology*, 27:89–99, 2001.

[92] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *Proceedings of 1st International Conference on Computer Vision, London, UK*, 1:259–268, 1987.

[93] D. G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. London Math. Soc.*, 16:81–121, 1984.

[94] D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4:87–99, 1989.

[95] S. Kim, Y.-H. Kim. Image flow segmentation and estimation using displaced spatialgradient. *IEEE Electronics Letters*, 28:2213–2215, 1992.

[96] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19:41–47, 1986.

[97] H. Knutsson and M. Andersson. Morphons: Paint on priors and elastic canvas for segmentation and registration. 3540:46–55, 2005.

[98] C. Kotropoulos and I. Pitas. Segmentation of ultrasonic images using Support Vector Machines. *Pattern Recognition Letters*, 24:715–727, 2003.

[99] K. Koutroumbas and S. Theodoridis. *Pattern Recognition*. 1998.

[100] V. Kovalev, N. Harder, B. Neumann, M. Held, U. Liebel, H. Erfle, J. Ellenberg, B. Neumann, R. Eils, and K. Rohr. Feature selection for evaluating fluorescence microscopy images in genome-wide cell screens. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:276–283, 2006.

[101] N. Kozic, M. A. G. Ballester, P. Büchler, N. Reimers, L. P. Nolte, M. G. Linguraru, and M. Reyes. Population-specific evaluation of implant bone

fitting using pca shape space and level sets. *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging*, 1:883–886, 2009.

[102] D. Kritchevsky. Atherosclerosis and cholesterol lipids and human health. *Dairy Sci.*, 50:776–81, 1967.

[103] S. Kudo, K. Hosoe, M. Hosobuchi, N. Kawasaki, and K. Tanishita. Atp release from cultured endothelial cells and intercellular calcium signaling during shear stress exposure. *Journal of Biomechanical Science and Engineering*, 4:274–285, 2009.

[104] C. Kuo-Yu, C. Shih-Che, and L. Sheng-Fuu. Chin contour extraction based on an auto-initialized shape-enhanced snake. *International Journal of Computer Science and Network Security*, 10:46–55, 2010.

[105] F. Lamberti and B. Montrucchio. Segmentation of *in-vitro* endothelial cell networks. *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, 1:129–132, 2004.

[106] O. V. Larsen, P. Radeva, and E. Marti. Bounds on the optimal elasticity parameters for a snake. *ICIAP*, 1:37–42, 1995.

[107] A. Law, K. Lam, F. Lam, T. Wong, J. Poon, and F. Chan. Image analysis system for assessment of immunohistochemically stained proliferative marker (mib-1) in oesophageal squamous cell carcinoma. *Computer Methods and Programs in Biomedicine*, 70:37–45, 2003.

[108] C.-C. Lee, Y.-C. Chiang, C.-L. Tsai, and S.-H. Chen. Distinction of liver disease from CT images using kernel-based classifiers. *ICMED*, 1:113–120, 2007.

[109] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner. Segmenting brain tumors with conditional random fields and support vector machines. *Computer Vision for Biomedical Image Applications*, 3765:469–478, 2005.

[110] B. Lenseigne, P. Brodin, J. H. Kyoung, T. Christophe, and A. Genovesio. Support vector machines for automatic detection of tuberculosis bacteria in confocal microscopy images. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1:85–88, 2007.

[111] M. J. Levesque and R. M. Nerem. The elongation and orientation of cultured endothelial cells in response to shear stress. *Journal of Biomechanical Engineering*, 107:341–347, 1985.

[112] F. Leymarie and M. Levine. Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:617–634, 2002.

[113] P. Libby, P. M. Ridker, and A. Maseri. Inflammation and atherosclerosis. *Circulation*, 105:1135–1143, 2002.

[114] Y. Lin and K. Fu. Automatic classification of cervical cells using a binary tree classifier. *Pattern Recognition*, 16:69–80, 1983.

[115] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31:983–1001, 1998.

[116] C. Loukas, G. D. Wilson, and B. Vojnovic. Automated segmentation of cancer cell nuclei in complex tissue sections. *Proc. SPIE*, 4158:188–198, 2001.

[117] C. G. Loukas and A. Linney. A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. *Computer Methods and Programs in Biomedicine*, 74:183–199, 2004.

[118] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

[119] A. M. Malek, S. L. Alper, and S. Izumo. Hemodynamic shear stress and its role in atherosclerosis. *JAMA: The Journal of the American Medical Association*, 282:2035–2042, 1999.

Bibliography

[120] J. I. Markus, M. Peura, J. Srel, and A. Visa. Comparison of combined shape descriptors for irregular objects. *In Proceedings of the Eight British Machine Vision Conference*, pages 430–439, 1997.

[121] T. MathWorks. Matlab, Version 7.9.0 (R2009b) on PC Windows. http://www.mathworks.co.uk/.

[122] U. Maulik and S. Bandyopadhyay. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1650–1654, 2002.

[123] B. W. McCrindle, E. M. Urbina, B. A. Dennison, M. S. Jacobson, J. Steinberger, A. P. Rocchini, L. L. Hayman, and S. R. Daniels. Drug therapy of high-risk lipid abnormalities in children and adolescents: A scientific statement from the American Heart Association atherosclerosis, hypertension, and obesity in youth committee, council of cardiovascular disease in the young, with the council on cardiovascular nursing. *Circulation*, 115:1948–1967, 2007.

[124] U. D. Meester, I. T. Young, J. Lindeman, and H. C. van. Der. Linden. Towards a quantitative grading of bladder tumors. *Cytometry*, 12:602–613, 1991.

[125] L. R. Monteiro, J. A. F. Diniz-Filho, S. F. dos Reis, and E. D. Arajo. Geometric estimates of heritability in biological shape. *Evolution*, 56:563–572, 2002.

[126] T. Mouroutis, S. Roberts, A. Bharath, and G. Alusi. Compact hough transform and a maximum likelihood approach to cell nuclei detection. *13th International Conference on Digital Signal Processing Proceedings*, 2:869–872, 1997.

[127] R. F. Murphy, M. Velliste, and G. Porreca. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. *The Journal of VLSI Signal Processing*, 35:311–321, 2003.

[128] W. Niblack. *An introduction to digital image processing.* Strandberg Publishing Company, 1985.

[129] S. Nixon and A. S. Aguado. *Feature Extraction and Image Processing,* volume Second Edition,. 2nd ed edition, 2008.

[130] M. U. Nollert, S. L. Diamond, and L. V. McIntire. Hydrodynamic shear stress and mass transport modulation of endothelial cell metabolism. *Biotechnology and bioengineering*, 38:588–602, 1991.

[131] S. H. Ong, X. C. Jin, and R. Jayasooriah, Sinniah. Image analysis of tissue sections. *Computers in Biology and Medicine*, 26:269–279, 1996.

[132] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[133] U. Pal, K. Rodenacker, and B. Chaudhuri. Automatic cell segmentation in cyto- and histometry using dominant contour feature points. *Analytical Cellular Pathology*, 17:243–250, 1998.

[134] G. Pasterkamp and E. Falk. Atherosclerotic plaque rupture: an overview. *Journal of clinical and basic cardiology*, 3:81–86, 2000.

[135] T. Pavlidis. Algorithms for shape analysis of contours and waveforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:301 –312, 1980.

[136] P. Pedersen, V. Mitra, and J. Dey. Boundary detection in 3D ultrasound reconstruction using nearest neighbor map. *Proc. SPIE*, 6147:57–67, 2006.

[137] E. Persoon and K.-S. Fu. Shape discrimination using fourier descriptors. *IEEE Transactions on Systems, Man and Cybernetics*, 7:170–179, 1977.

[138] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[139] J. A. F. Pierna, V. Baeten, A. M. Renier, R. P. Cogdill, and P. Dardenne. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Chemometrics*, 18:341–349, 2004.

[140] Z. Pincus and J. A. Theriot. Comparison of quantitative methods for cell-shape analysis. *Journal of Microscopy*, 227:140–156, 2007.

[141] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney. Deformable for 3D medical image segmentation. *International Journal of Computer Vision*, 55:85–106, 2003.

[142] N. Poulter. Global risk of cardiovascular disease. *Heart*, 89:2–5, 2003.

[143] S. Rahnamayan, H. Tizhoosh, and M. Salama. Automated snake initialization for the segmentation of the prostate in ultrasound images. *Image Analysis and Recognition*, 3656:930–937, 2005.

[144] J. Rajapakse, J. Giedd, and J. Rapoport. Statistical approach to segmentation of single-channel cerebral mr images. *IEEE Transactions on Medical Imaging*, 16:176 –186, 1997.

[145] W. Reddick, J. Glass, E. Cook, T. Elkin, and R. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on Medical Imaging*, 16:911–918, 1997.

[146] P. L. Rosin. Unimodal thresholding. *Pattern Recognition*, 34:2083–2096, 2001.

[147] R. Ross. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature*, 362:801–809, 1993.

[148] P. K. Sahoo, S. Soltani, A. K. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41:233–260, 1988.

[149] S. C. Sekhar, F. Aguet, S. Romain, P. Thevenaz, and M. Unser. Parametric B-spline snakes on distance maps-application to segmentation of histology images. *16th European Signal Processing Conference Lausanne, Switzerland*, 2008.

[150] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13:146–165, 2004.

[151] J. Shawe-Taylor and N. Cristianini. Support Vector Machines and other kernel-based learning methods. *Cambridge University Press*, 2000.

[152] J. Shen, F. W. Luscinskas, A. Connolly, J. Dewey, C. F., and J. Gimbrone, M. A. Fluid shear stress modulates cytosolic free calcium in vascular endothelial cells. *Am J Physiol Cell Physiol*, 262:384–390, 1992.

[153] F. Y. Shih and S. Cheng. Adaptive mathematical morphology for edge linking. *Information Sciences*, 167:9–21, 2004.

[154] Y. Shiogai, K. H. Parker, K. Kitamura, and S. R. Schultz. A simple method for extracting spike trains from two-photon calcium imaging data, Personal Communication.

[155] A. W. Sinzinger H, Silberbauer K. Quantitative investigation of sudanophilic lesions around the aortic ostia of human fetuses, newborn and children. *Blood Vessels*, 17:44–52, 1980.

[156] P. J. Sjstrm, B. R. Frydel, and L. U. Wahlberg. Artificial neural network-aided image analysis system for cell counting. *Cytometry*, 36:18–26, 1999.

[157] C. Sommer. Quantitative characterization, classification and reconstruction of oocyst shapes of eimeria species from cattle. *Parasitology*, 116:21–28, 1998.

[158] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

Bibliography

[159] H. C. Stary. Natural history and histological classification of atherosclerotic lesions: an update. *Arterioscler Thromb Vasc Biol.*, 20, 2000.

[160] T. Stevens, S. Phan, M. G. Frid, D. Alvarez, E. Herzog, and K. R. Stenmark. Lung vascular cell heterogeneity: Endothelium, smooth muscle, and fibroblasts. *Proc Am Thorac Soc*, 5:783–791, 2008.

[161] M. R. Teague. Image analysis via the general theory of moments. *J. Opt. Soc. Am.*, 70:920–930, 1980.

[162] N. Theera-Umpon and S. Dhompongsa. Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. *Information Technology in Biomedicine, IEEE Transactions on*, 11:353–359, 2007.

[163] J.-P. Thiran and B. Macq. Morphological feature extraction for the classification of digital images of cancerous tissues. *IEEE Transactions on Biomedical Engineering*, 43:1011–1020, 1996.

[164] A. Thomas, T. Davies, and A. Luxmoore. The hough transform for locating cell nuclei. *IEE Colloquium on Applications of Image Processing in Mass Health Screening*, 8:1–4, 1992.

[165] C. E. Todman A.G. Low-level grouping mechanisms for contour completion. *Information Sciences*, 125:19–35, 2000.

[166] I. Toma, E. Bansal, E. J. Meer, J. J. Kang, S. L. Vargas, and J. Peti-Peterdi. Connexin 40 and ATP-dependent intercellular calcium wave in renal glomerular endothelial cells. *Am J Physiol Regul Integr Comp Physiol*, 294:1769–1776, 2008.

[167] H. Trichili, M.-S. Bouhlel, N. Derbel, and L. Kamoun. A survey and evaluation of edge detection operators application to medical images. *IEEE International Conference on Systems, Man and Cybernetics*, 4:4, 2002.

[168] M. Tscherepanow, F. Zollner, and F. Kummert. Classification of segmented regions in brightfield microscope images. *18th International Conference on Pattern Recognition*, 3:972–975, 2006.

[169] R. Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry*, 67:509544, 1998.

[170] A. P. Umesh and B. Chaudhuri. Region based techniques for segmentation of volumetric histo-pathological images. *Computer Methods and Programs in Biomedicine*, 61:23–47, 2000.

[171] V. Vapnik. The nature of statistical learning theory. *Springer-Verlag*, 1995.

[172] M. Velliste and R. Murphy. Automated determination of protein subcellular locations from 3D fluorescence microscope images. *IEEE International Symposium on Biomedical Imaging*, 1:867–870, 2002.

[173] D. L. Vilario, V. M. Brea, D. Cabello, and J. M. Pardo. Discrete-time cnn for image segmentation by active contours. *Pattern Recognition Letters*, 19:721–734, 1998.

[174] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:583–598, 1991.

[175] L. M. Vincent and B. R. Masters. Morphological image processing and network analysis of cornea endothelial cell images. *Proc. SPIE*, 1769, 1992.

[176] R. Walker, P. Jackway, B. Lovell, and I. D. Longstaff. Classification of cervical cell nuclei using morphological segmentation and textural feature extraction. *Proceedings of Second Australian and New Zealand Conference on Intelligent Information Systems*, 1:297–301, 1994.

[177] M. Wang, X. Zhou, F. Li, J. Huckins, R. W. King, and S. T. Wong. Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics*, 24:94–101, 2008.

[178] J. R. Weaver and J. L.-S. Au. Application of automatic thresholding in image analysis scoring of cells in human solid tumors labeled for proliferation markers. *Cytometry*, 29:128–135, 1997.

[179] L. Wei, L. Y. sheng, R. X. yi, and D. Peng. A new contour detection in mammogram using sequential edge linking. *Second International Symposium on Intelligent Information Technology Application*, 1:197–200, 2008.

[180] P. D. Weinberg. Disease patterns at arterial branches and their relation to flow. *Biorheology*, 39:533–537, 2002.

[181] H.-S. Wu, J. Barba, and J. Gil. A parametric fitting algorithm for segmentation of cell images. *IEEE Transactions on Biomedical Engineering*, 45:400–407, 1998.

[182] H.-S. Wu, J. Berba, and J. Gil. Iterative thresholding for segmentation of cells from noisy images. *Journal of Microscopy*, 197:296–304, 2000.

[183] P. Wust, J. Gellermann, J. Beier, S. Wegner, W. Tilly, J. Trger, D. Stalling, H. Oswald, H. C. Hege, P. Deuflhard, and R. Felix. Evaluation of segmentation algorithms for generation of patient models in radiofrequency hyperthermia. *Physics in Medicine and Biology*, 43:32953307, 1998.

[184] C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7:359–369, 1998.

[185] J. C. Y.-L. Fok and R. Chin. Automated analysis of nerve-cell images using active contour models. *IEEE Trans. Med. Imaging*, 15:353–368, 1996.

[186] H. Yamada, C. Merritt, and T. Kasvand. Recognition of kidney glomerulus by dynamic programming matching method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10:731–737, 1988.

[187] K. Yamamoto. Impaired flow-dependent control of vascular tone and remodeling, *in p2x4-deficient mice. Nature Medicine*, 12:133, 2005.

[188] K. Yamamoto, T. Sokabe, N. Ohura, H. Nakatsuka, A. Kamiya, and J. Ando. Endogenously released ATP mediates shear stress-induced Ca2+ influx into pulmonary artery endothelial cells. *Am J Physiol Heart Circ Physiol*, 285:793–803, 2003.

Bibliography

[189] Y. S. Yang, D. K. Park, H. C. Kim, M.-H. Choi, and J.-Y. Chai. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Transactions on Biomedical Engineering*, 48:718–730, 2001.

[190] M. L. Zelditch, D. L. Swiderski, D. H. Sheets, and W. L. Fink. *Geometric Morphometrics for Biologists*. Academic Press, 2004.

[191] H. Zhang, J. E. Fritts, and S. A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110:260–280, 2008.

[192] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29:1335–1346, 1996.

[193] Y. J. Zhang. A review of recent evaluation methods for image segmentation. *Sixth International Symposium on.Signal Processing and its Applications*, 1:148–151, 2001.

[194] S. Zhao, A. Suciu, T. Ziegler, J. Moore, James E., E. Burki, J.-J. Meister, and H. R. Brunner. Synergistic Effects of Fluid Shear Stress and Cyclic Circumferential Stretch on Vascular Endothelial Cell Morphology and Cytoskeleton. *Arterioscler Thromb Vasc Biol*, 15:1781–1786, 1995.

[195] P. Zhou and D. Pycock. Robust statistical models for cell image interpretation. *Image and Vision Computing*, 15:307–316, 1997.

[196] G. Zinser and D. Komitowski. Segmentation of cell nuclei in tissue section analysis. *J. Histochem. Cytochem.*, 31:94–100, 1983.

A. Appendix

A.1. Some Results of Chapter 4
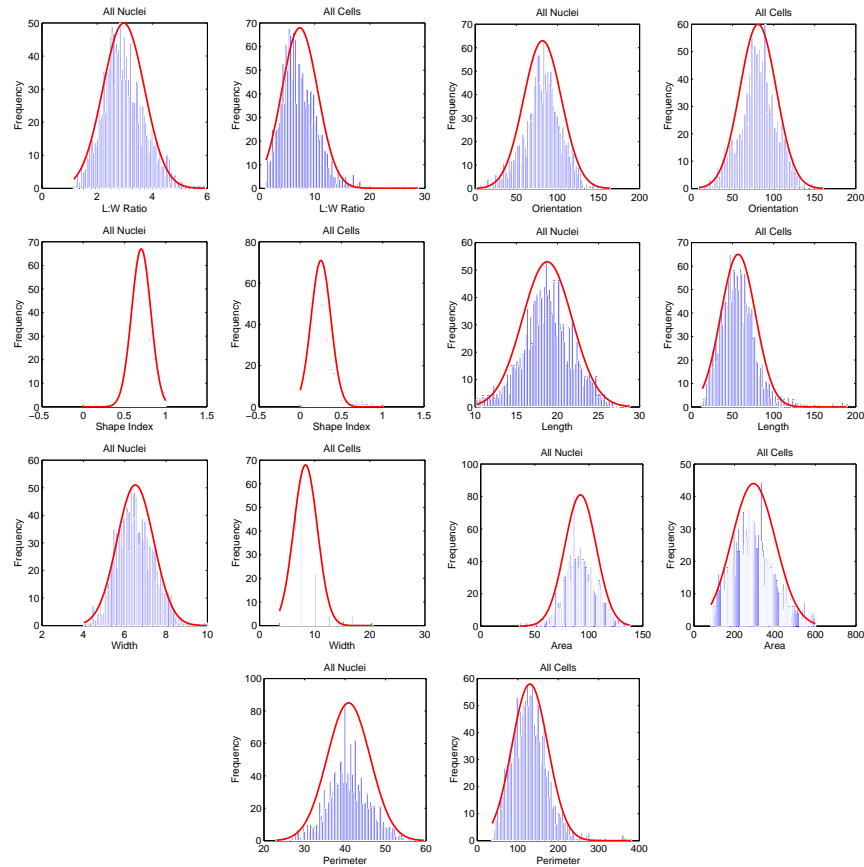
Figure A.1.: Histograms and normal distribution, as indicated by the superimposed curve, of different shape parameters for all endothelial cells and nuclei, here the total number of cells are 1496.

Figure A.2.: Cells versus nuclei mean $\pm$ STD, mean $\pm$ SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of ratios for mature and immature cells, here total number of cells are 559 and 937 respectively.

Figure A.3.: Cells versus nuclei mean ± STD, mean ± SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of orientations for mature and immature cells, here total number of cells are 559 and 937 respectively.

Figure A.4.: Cells versus nuclei mean ± STD, mean ± SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of shape index for mature and immature cells, here total number of cells are 559 and 937 respectively.

Figure A.5.: Cells versus nuclei mean ± STD, mean ± SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of lengths for mature and immature cells, here total number of cells are 559 and 937 respectively.

Figure A.6.: Cells versus nuclei mean ± STD, mean ± SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of widths for mature and immature cells, here total number of cells are 559 and 937 respectively.

Figure A.7.: Cells versus nuclei mean ± STD, mean ± SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of areas for mature and immature cells, here total number of cells are 559 and 937 respectively.

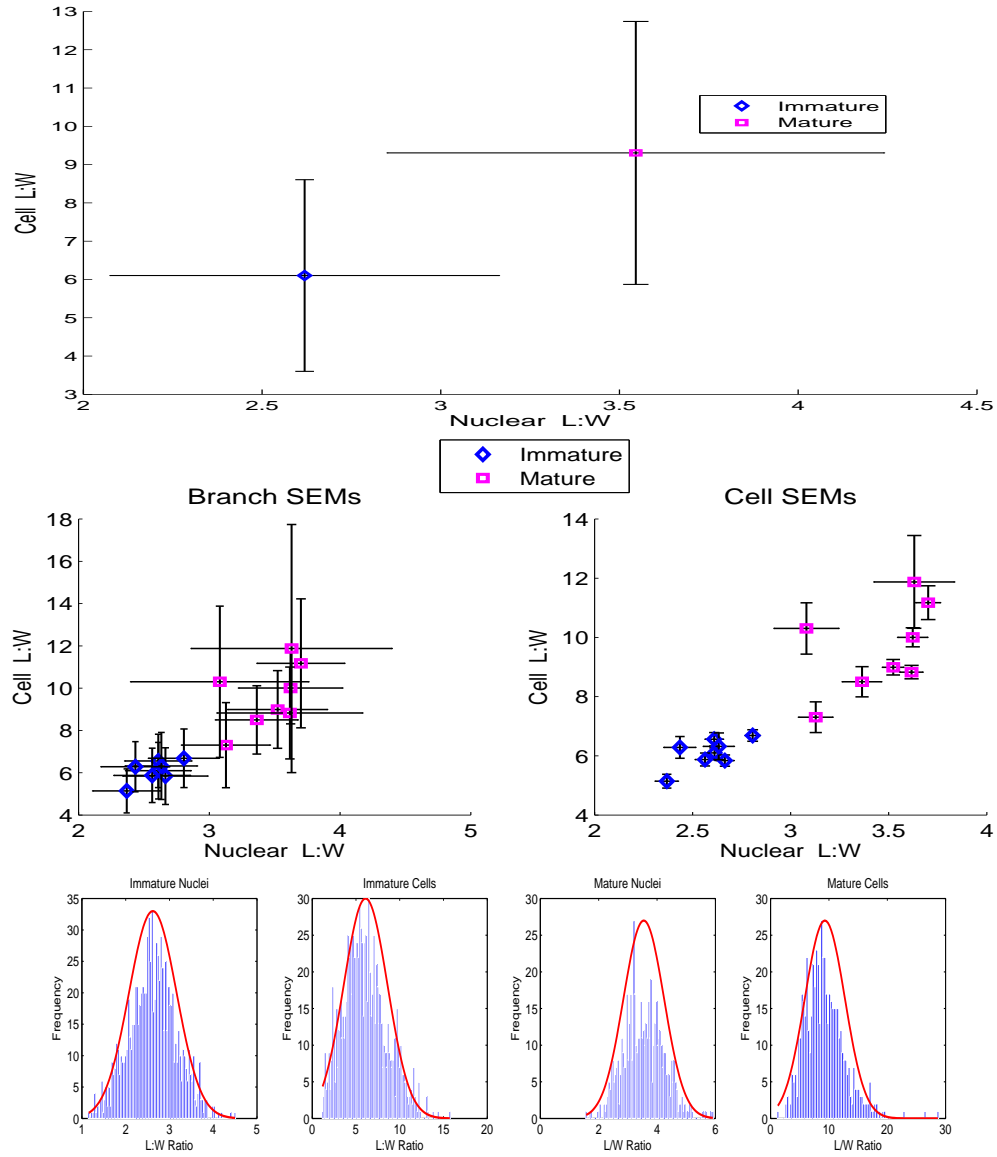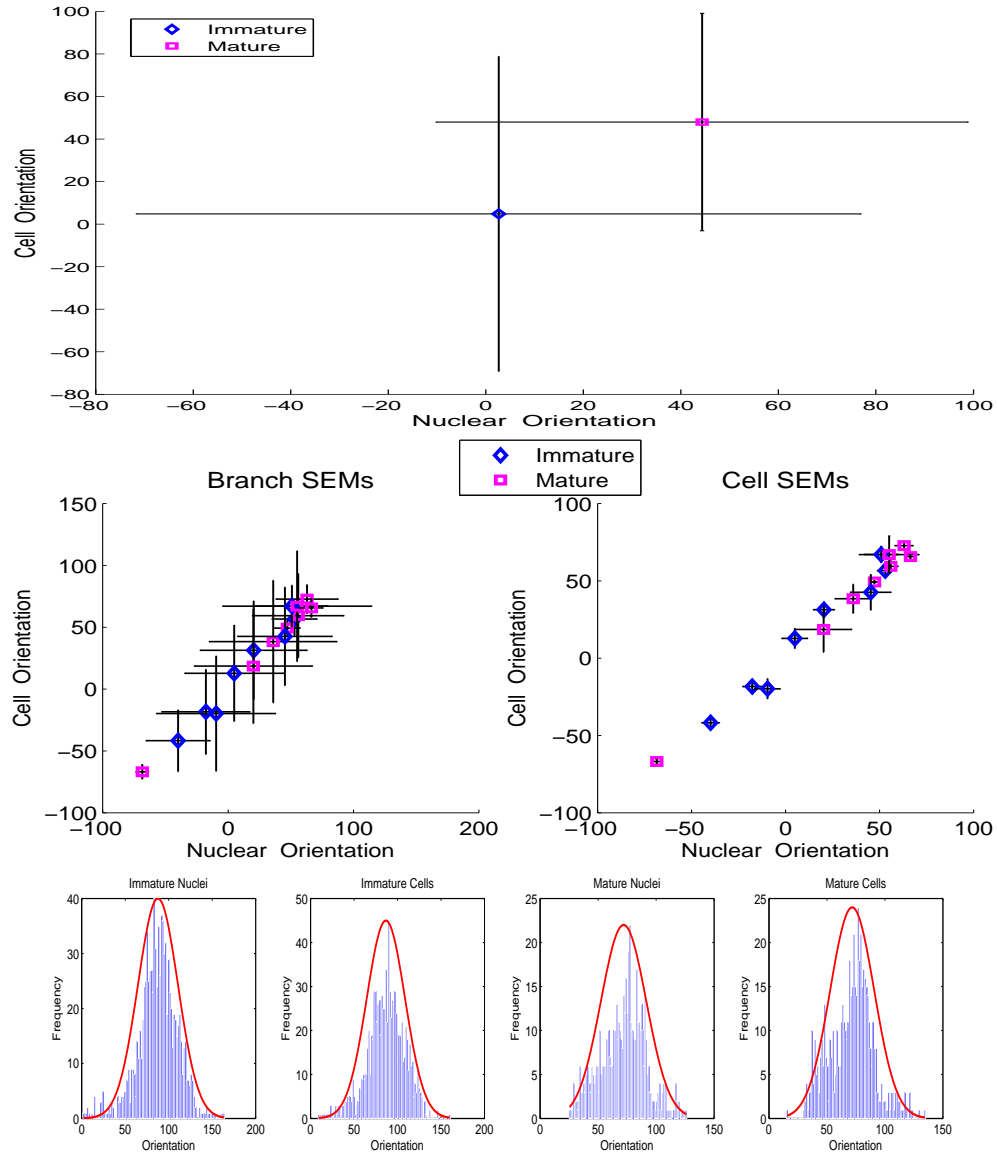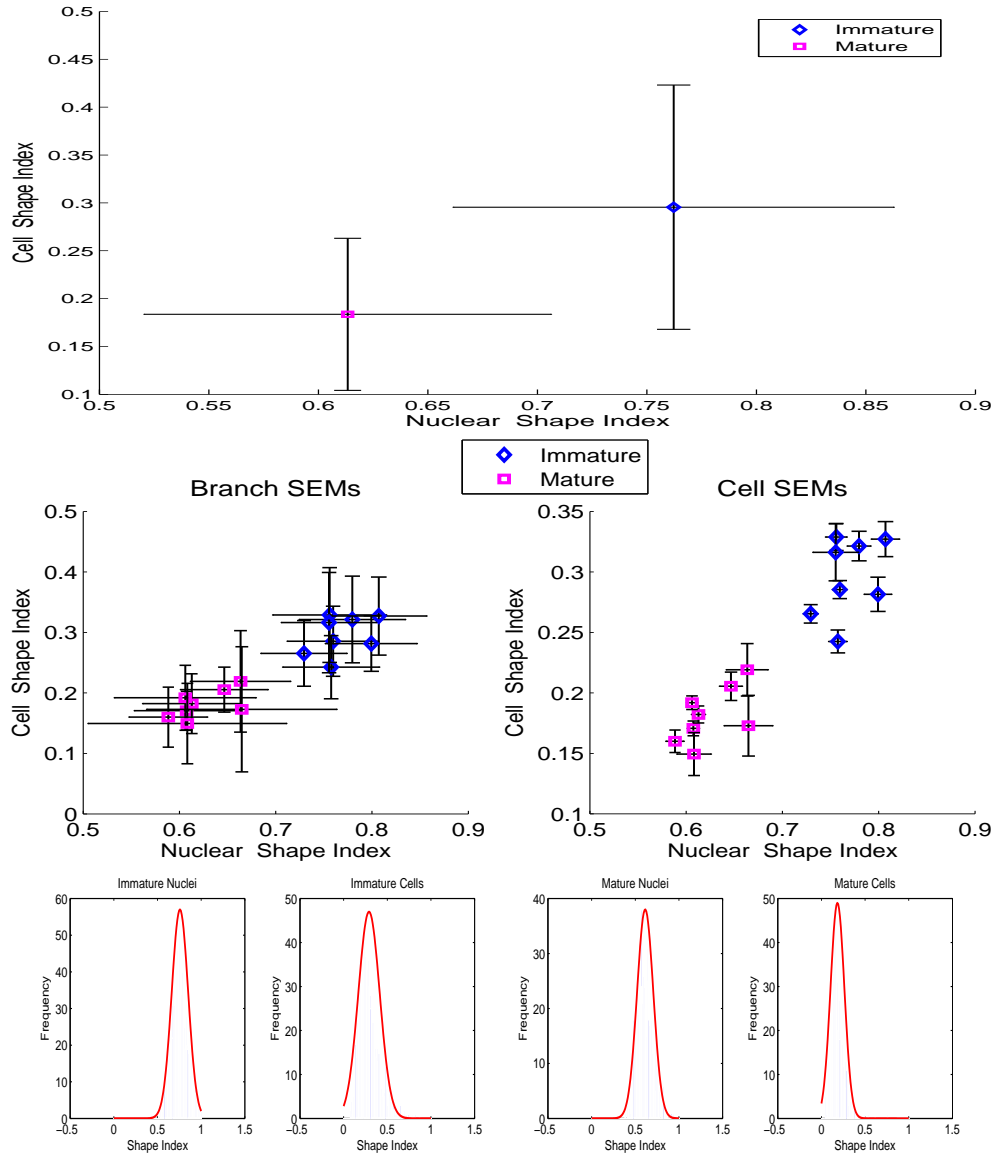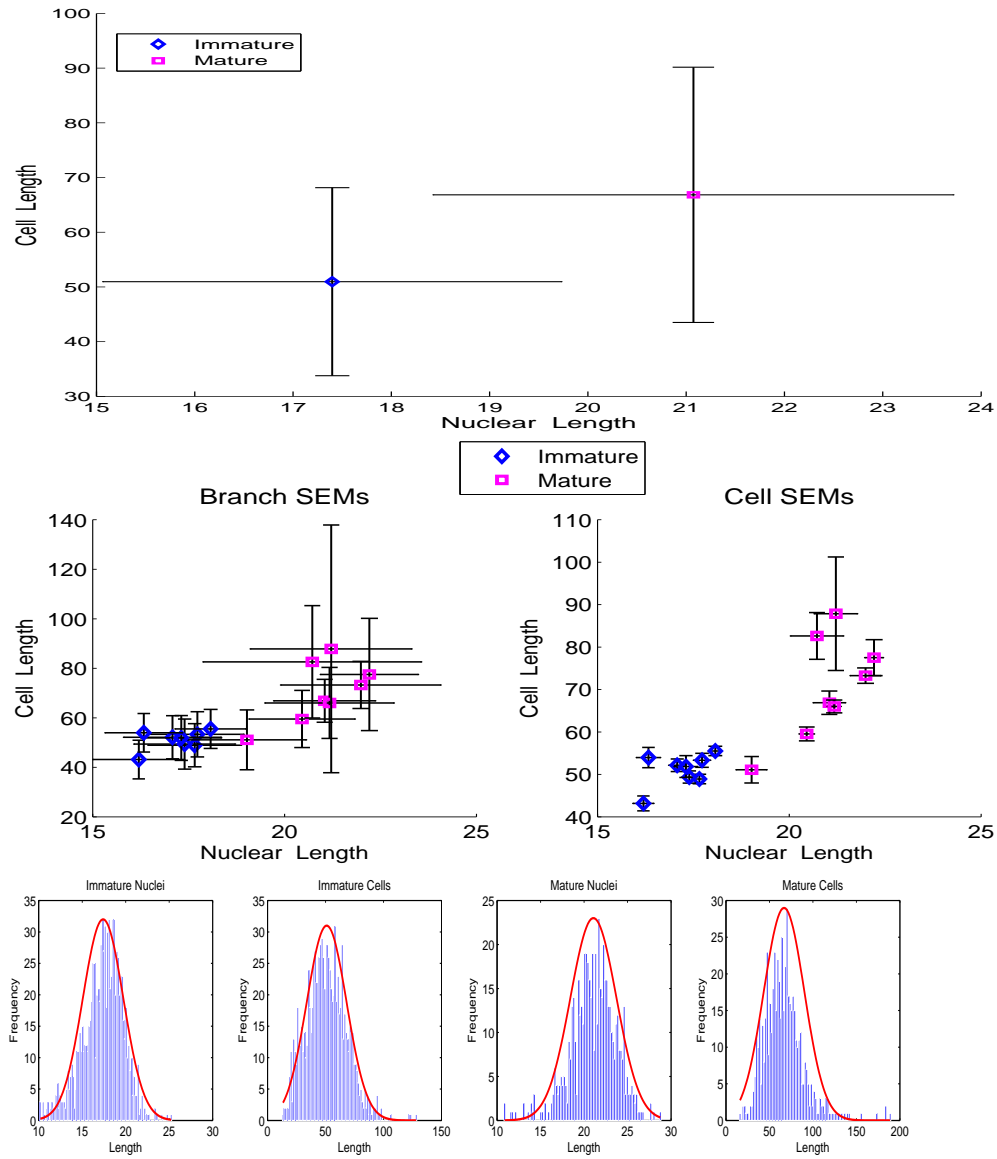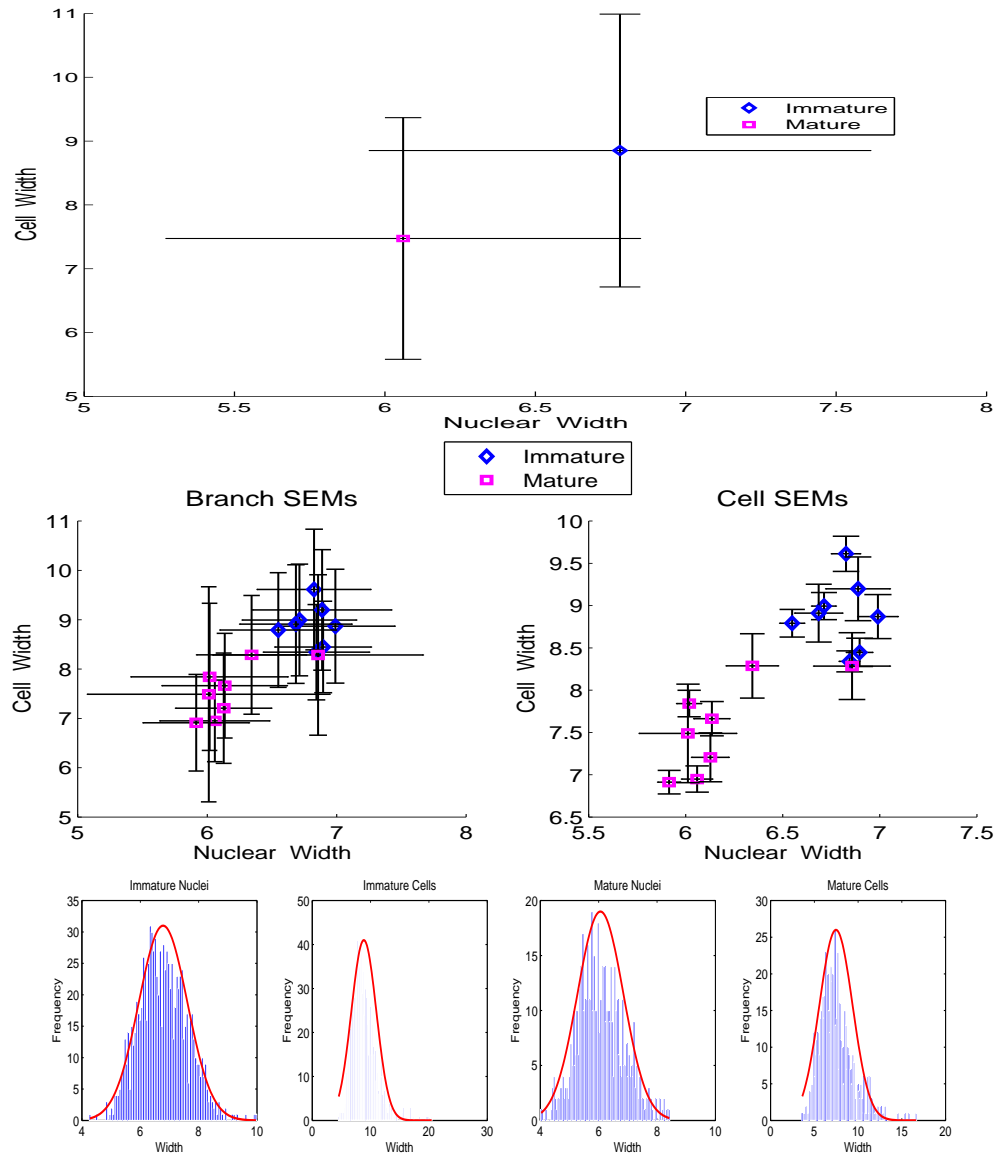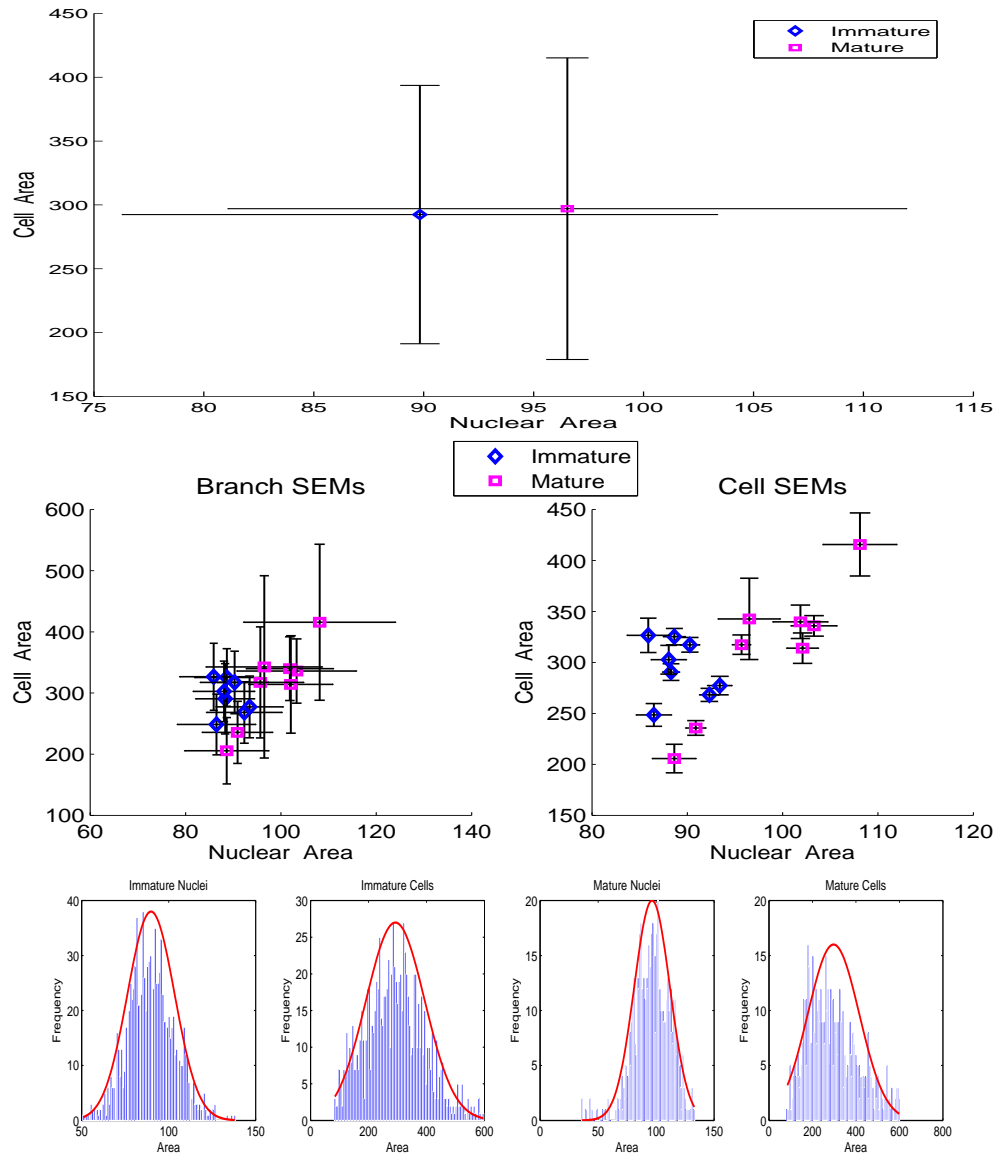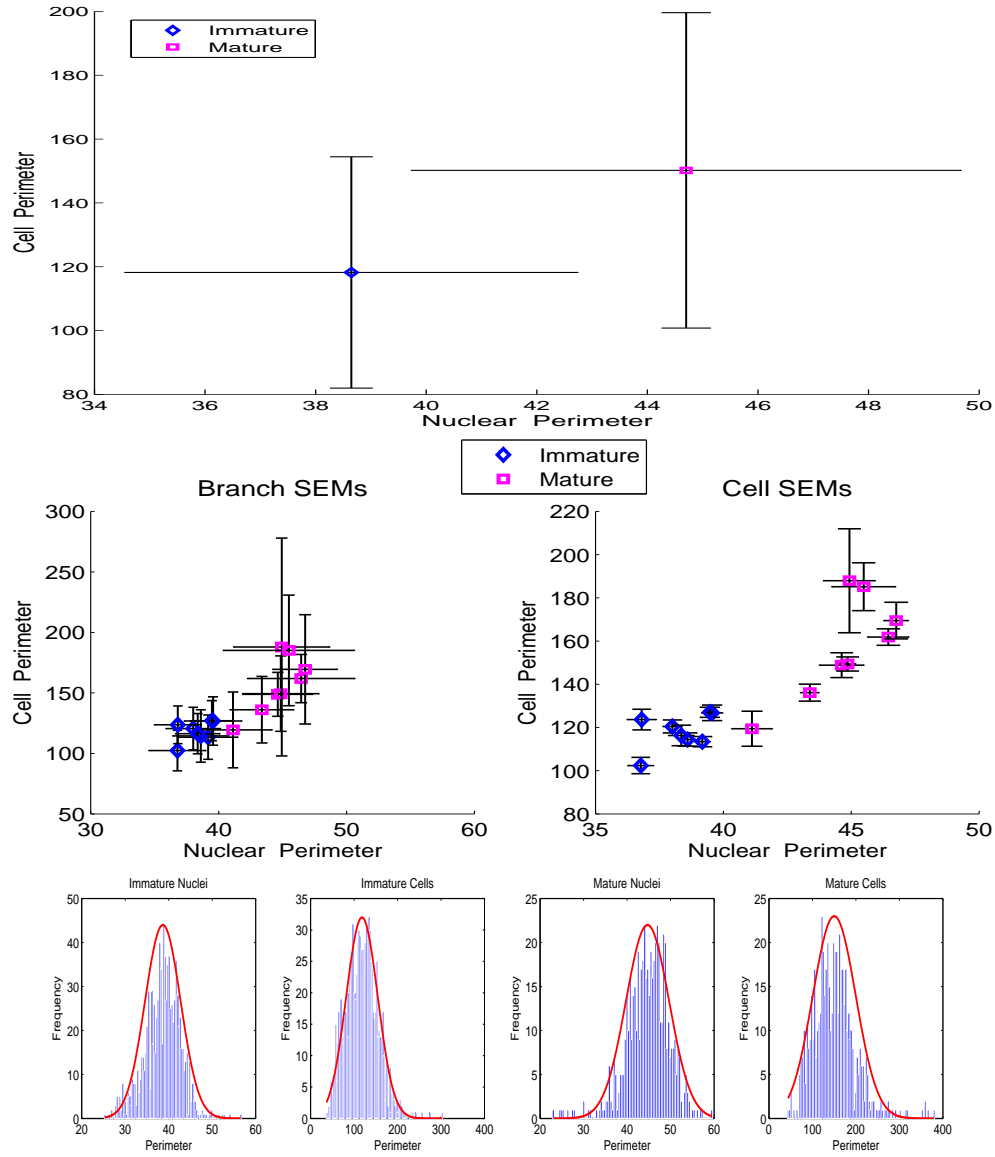Figure A.8.: Cells versus nuclei mean $\pm$ STD, mean $\pm$ SEMs over branch and over cell and histograms and normal distribution, as indicated by the superimposed curve, of perimeters for mature and immature cells, here total number of cells are 559 and 937 respectively.

		C #	L:W Ratio			Shape Index			Angle (°)			Length (μm)		
			m	σ/Br	σ/C	m	σ/Br	σ/C	m	σ/Br	σ	m	σ/Br	σ/C
UL	EC	264	7.31	0.19	1.27	0.25	0.01	0.04	44.22	3.78	25.09	56.48	1.18	7.83
	ECn	264	2.93	0.04	0.29	0.71	0.01	0.05	33.66	4.20	27.86	19.00	0.21	1.36
U	EC	158	6.94	0.22	1.01	0.24	0.01	0.04	-29.91	4.97	23.24	55.98	1.49	6.97
	ECn	158	2.76	0.05	0.23	0.74	0.01	0.04	-28.06	5.05	23.60	18.38	0.22	1.04
UR	EC	256	7.03	0.24	1.56	0.29	0.01	0.06	-5.70	4.55	29.75	55.65	1.63	10.66
	ECn	256	2.90	0.05	0.31	0.72	0.01	0.05	-5.77	4.60	30.06	18.47	0.20	1.30
L	EC	168	6.13	0.21	1.09	0.30	0.01	0.06	-1.40	5.96	31.52	51.97	1.34	7.07
	ECn	168	2.66	0.05	0.25	0.76	0.01	0.05	-8.21	5.90	31.21	17.42	0.20	1.06
R	EC	227	7.65	0.22	1.28	0.23	0.01	0.05	51.86	1.21	6.90	53.83	1.33	7.60
	ECn	227	3.12	0.05	0.31	0.68	0.01	0.05	49.25	1.67	9.53	18.97	0.21	1.18
DL	EC	219	8.41	0.21	1.41	0.21	0.01	0.05	56.54	3.51	23.26	63.62	1.37	9.07
	ECn	219	3.45	0.06	0.37	0.63	0.01	0.05	54.11	3.64	24.06	20.52	0.18	1.21
D	EC	56	7.68	0.57	1.91	0.25	0.01	0.05	67.10	4.79	16.04	62.44	4.21	14.08
	ECn	56	2.73	0.10	0.35	0.75	0.01	0.06	51.93	7.66	25.64	17.55	0.39	1.31
DR	EC	168	7.05	0.21	1.38	0.26	0.01	0.05	-11.12	6.25	40.49	58.26	1.30	8.42
	ECn	168	2.83	0.04	0.25	0.72	0.01	0.04	-1.91	6.24	40.45	18.34	0.15	0.96
EC ranges			6.13 - 8.41			0.21 - 0.30			-29.91 - 67.10			51.97 - 63.62		
ECn ranges			2.66 - 3.45			0.63 - 0.76			-28.06 - 54.11			17.42 - 20.52		

		C #	Width (μm)			Area (Pixels²)			Perimeter (Pixels)		
			m	σ/Br	σ/C	m	σ/Br	σ/C	m	σ/Br	σ
UL	EC	264	8.13	0.11	0.72	289.21	5.75	38.16	128.49	2.43	16.13
	ECn	264	6.62	0.05	0.33	95.73	1.06	7.05	41.44	0.38	2.53
U	EC	158	8.42	0.15	0.72	289.57	8.23	38.48	131.10	3.18	14.89
	ECn	158	6.79	0.06	0.29	95.07	1.19	5.57	40.55	0.40	1.87
UR	EC	256	8.54	0.14	0.91	295.74	7.17	46.81	126.69	3.35	21.91
	ECn	256	6.57	0.05	0.36	91.37	0.85	5.53	40.42	0.34	2.20
L	EC	168	9.18	0.19	1.02	303.93	7.88	41.68	120.29	2.85	15.06
	ECn	168	6.70	0.07	0.35	88.64	1.02	5.38	38.56	0.34	1.79
R	EC	227	7.59	0.14	0.81	240.19	6.11	34.81	124.39	3.06	17.42
	ECn	227	6.29	0.06	0.36	89.36	0.94	5.37	41.08	0.36	2.02
DL	EC	219	8.02	0.14	0.96	314.91	8.26	54.66	143.82	2.95	19.54
	ECn	219	6.13	0.06	0.39	94.40	0.87	5.73	43.76	0.32	2.10
D	EC	56	8.77	0.33	1.11	330.63	15.96	53.40	139.74	7.82	26.18
	ECn	56	6.67	0.15	0.49	88.55	1.92	6.42	38.84	0.68	2.27
DR	EC	168	8.74	0.15	0.99	327.24	7.52	48.71	132.93	2.73	17.68
	ECn	168	6.58	0.06	0.39	92.08	1.07	6.91	40.08	0.28	1.82
EC ranges			7.59 - 9.18			240.19 - 330.63			120.29 - 143.82		
ECn ranges			6.13 - 6.79			88.55 - 95.73			38.56 - 43.76		

Table A.1.: Vascular region-wise mean ± SEM over branch and over cell of EC and ECn morphology parameters for all cells. Here total number of cells is 1496, m = Mean , σ = SEM, C = cells and Br = Branches.

		C #	L:W Ratio m	σ/Br	σ/C	Shape Index m	σ/Br	σ/C	Angle (°) m	σ/Br	σ	Length (μm) m	σ/Br	σ/C
UL	EC	182	6.10	0.17	1.34	0.29	0.01	0.06	31.36	5.12	39.91	48.92	1.12	8.70
	ECn	182	2.61	0.03	0.25	0.76	0.01	0.05	20.44	5.44	42.38	17.66	0.16	1.22
U	EC	123	6.56	0.23	1.26	0.24	0.01	0.05	-41.73	4.47	24.76	53.32	1.65	9.13
	ECn	123	2.61	0.05	0.26	0.76	0.01	0.05	-39.83	4.58	25.38	17.73	0.22	1.22
UR	EC	199	5.84	0.19	1.34	0.33	0.01	0.08	-18.35	4.88	34.42	49.38	1.44	10.12
	ECn	199	2.66	0.05	0.32	0.76	0.01	0.06	-17.72	4.96	34.95	17.39	0.19	1.32
L	EC	138	5.88	0.22	1.29	0.32	0.01	0.07	12.83	6.66	39.09	52.16	1.48	8.71
	ECn	138	2.56	0.05	0.29	0.78	0.01	0.06	4.89	6.67	39.20	17.08	0.22	1.27
R	EC	79	5.14	0.23	1.04	0.33	0.01	0.06	56.64	2.99	13.30	43.15	1.76	7.80
	ECn	79	2.37	0.06	0.26	0.81	0.01	0.05	52.97	4.00	17.76	16.20	0.27	1.19
DL	EC	37	6.32	0.45	1.58	0.32	0.02	0.08	42.65	11.28	39.60	51.86	2.57	9.04
	ECn	37	2.63	0.08	0.28	0.76	0.02	0.05	45.29	10.67	37.46	17.30	0.30	1.05
D	EC	42	6.29	0.37	1.19	0.28	0.01	0.05	67.12	5.12	16.58	53.96	2.40	7.78
	ECn	42	2.44	0.08	0.26	0.80	0.01	0.05	50.87	8.86	28.70	16.33	0.31	1.01
DR	EC	151	6.69	0.20	1.39	0.27	0.01	0.05	-19.77	6.59	46.73	55.52	1.11	7.87
	ECn	151	2.81	0.04	0.27	0.73	0.01	0.04	-9.60	6.66	47.23	18.07	0.13	0.92
EC ranges			5.14 - 6.69			0.24 - 0.33			-41.73 - 67.12			43.15 - 55.52		
ECn ranges			2.37 - 2.81			0.73 - 0.81			-39.83 - 52.97			16.20 - 18.07		

		C #	Width (μm) m	σ/Br	σ/C	Area (Pixels²) m	σ/Br	σ/C	Perimeter (Pixels) m	σ/Br	σ
UL	EC	182	8.34	0.12	0.97	268.19	6.44	50.19	113.44	2.35	18.28
	ECn	182	6.84	0.05	0.41	92.32	1.01	7.89	39.18	0.30	2.31
U	EC	123	8.45	0.17	0.93	277.30	9.11	50.52	126.77	3.61	20.00
	ECn	123	6.90	0.07	0.37	93.41	1.27	7.02	39.56	0.40	2.24
UR	EC	199	8.99	0.16	1.13	290.51	8.15	57.49	114.43	3.07	21.66
	ECn	199	6.71	0.06	0.44	88.30	0.87	6.13	38.60	0.31	2.18
L	EC	138	9.61	0.21	1.22	325.29	8.02	47.14	120.48	3.00	17.62
	ECn	138	6.83	0.07	0.44	88.64	1.14	6.69	38.01	0.36	2.11
R	EC	79	8.87	0.26	1.15	248.54	11.18	49.68	102.38	3.77	16.77
	ECn	79	6.99	0.10	0.46	86.49	1.84	8.17	36.77	0.50	2.21
DL	EC	37	8.91	0.34	1.20	302.68	14.10	49.51	116.31	4.75	16.69
	ECn	37	6.69	0.12	0.43	88.05	1.82	6.41	38.36	0.49	1.73
D	EC	42	9.20	0.38	1.22	326.60	16.91	54.80	123.67	4.80	15.56
	ECn	42	6.89	0.17	0.54	85.90	2.19	7.09	36.81	0.56	1.82
DR	EC	151	8.79	0.16	1.16	317.28	7.21	51.13	127.05	2.34	16.59
	ECn	151	6.55	0.06	0.45	90.27	1.01	7.15	39.47	0.23	1.66
EC ranges			8.34 - 9.61			248.54 - 326.60			102.38 - 127.05		
ECn ranges			6.55 - 6.99			85.90 - 93.41			36.77 - 39.56		

Table A.2.: Vascular region-wise mean ± SEM over branch and over cell of EC and ECn morphology parameters for immature cells. Here total number of cells is 937, m = Mean , σ = SEM, C = cells and Br = Branches.

		C #	L:W Ratio m	σ/Br	σ/C	Shape Index m	σ/Br	σ/C	Angle (°) m	σ/Br	σ/C	Length (μm) m	σ/Br	σ/C
UL	EC	82	10.01	0.32	1.69	0.17	0.01	0.03	72.77	2.17	11.34	73.27	1.82	9.51
	ECn	82	3.62	0.08	0.40	0.61	0.01	0.05	63.00	4.70	24.60	21.99	0.40	2.09
U	EC	30	8.50	0.51	1.61	0.21	0.01	0.04	18.58	14.68	46.43	66.88	2.74	8.68
	ECn	30	3.36	0.10	0.32	0.65	0.01	0.05	20.21	14.84	46.92	21.05	0.42	1.33
UR	EC	57	11.18	0.57	3.05	0.16	0.01	0.05	38.44	9.24	49.32	77.52	4.24	22.66
	ECn	57	3.70	0.06	0.34	0.59	0.01	0.04	35.97	9.49	50.65	22.21	0.24	1.27
L	EC	30	7.31	0.52	2.01	0.22	0.02	0.08	-66.82	1.45	5.62	51.09	3.12	12.09
	ECn	30	3.13	0.09	0.34	0.66	0.02	0.05	-68.46	1.27	4.92	19.02	0.40	1.55
R	EC	148	8.99	0.26	1.84	0.18	0.01	0.05	49.30	0.89	6.27	59.53	1.64	11.52
	ECn	148	3.52	0.05	0.38	0.61	0.01	0.05	47.27	1.42	9.95	20.46	0.20	1.38
DL	EC	182	8.83	0.23	2.17	0.19	0.01	0.05	59.36	3.54	33.73	66.01	1.51	14.37
	ECn	182	3.62	0.06	0.56	0.61	0.01	0.07	55.91	3.80	36.26	21.17	0.18	1.68
D	EC	14	11.88	1.57	5.87	0.15	0.02	0.07	67.05	11.88	44.46	87.86	13.37	50.04
	ECn	14	3.63	0.21	0.77	0.61	0.02	0.10	55.12	15.81	59.16	21.21	0.56	2.10
DR	EC	17	10.30	0.87	3.58	0.17	0.03	0.10	65.69	1.78	7.33	82.62	5.51	22.71
	ECn	17	3.08	0.16	0.68	0.66	0.03	0.10	66.42	2.21	9.13	20.72	0.69	2.84
EC ranges			7.31 - 11.88			0.15 -0.22			-66.82 - 72.77			51.09 - 87.86		
ECn ranges			3.08 - 3.70			0.59 - 0.66			-68.46 - 66.42			19.02 - 22.21		

		C #	Width (μm) m	σ/Br	σ/C	Area (Pixels²) m	σ/Br	σ/C	Perimeter (Pixels) m	σ/Br	σ/C
UL	EC	82	7.66	0.20	1.06	335.87	10.05	52.55	161.89	3.81	19.92
	ECn	82	6.14	0.09	0.48	103.29	2.39	12.50	46.45	0.80	4.18
U	EC	30	8.29	0.38	1.20	339.87	16.45	52.03	148.87	5.75	18.18
	ECn	30	6.34	0.13	0.42	101.87	2.88	9.09	44.63	0.86	2.72
UR	EC	57	6.95	0.16	0.83	314.02	14.91	79.62	169.49	8.46	45.18
	ECn	57	6.06	0.08	0.42	102.09	1.65	8.79	46.76	0.47	2.51
L	EC	30	7.21	0.29	1.12	205.67	14.02	54.29	119.44	8.10	31.37
	ECn	30	6.13	0.10	0.37	88.63	2.28	8.81	41.12	0.79	3.05
R	EC	148	6.91	0.14	0.98	235.72	7.23	50.80	136.14	3.91	27.49
	ECn	148	5.91	0.06	0.41	90.89	1.05	7.34	43.38	0.35	2.48
DL	EC	182	7.84	0.16	1.49	317.40	9.52	90.81	149.42	3.27	31.19
	ECn	182	6.02	0.06	0.61	95.69	0.95	9.04	44.86	0.31	2.97
D	EC	14	7.49	0.58	2.18	342.71	39.85	149.09	187.94	24.06	90.02
	ECn	14	6.01	0.25	0.94	96.50	3.25	12.15	44.93	1.00	3.74
DR	EC	17	8.28	0.39	1.63	415.71	30.94	127.58	185.17	11.08	45.70
	ECn	17	6.86	0.20	0.81	108.12	3.85	15.89	45.49	1.24	5.11
EC ranges			6.91 - 8.29			205.67 - 415.71			119.44 - 187.94		
ECn ranges			5.91 - 6.86			88.63 - 108.12			41.12 - 46.76		

Table A.3.: Vascular region-wise mean ± SEM over branch and over cell of EC and ECn morphology parameters for mature cells. Here total number of cells is 559, m = Mean , σ = SEM, C = cells and Br = Branches.