Automated Quantitative and Qualitative Analysis of Neuroblastoma Cancer Tissue

Siamak Tafavogh

Doctor of Philosophy
University of Technology Sydney
2014

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(**Siamak Tafavogh**)

To My Beloved Parents

Acknowledgements

I would like to express my sincere gratitude to my PhD principle supervisor, Associate Prof. Paul Kennedy for all the effort that he put in during my PhD. During the PhD, there were critical times which would not have been overcome without his unsparing help. His professional, academic attitude, sense of humour, conscientious working style and generous personality will be of great benefit to me in my future research work and life.

I also would like to thank my co-supervisor Dr. Daniel Catchpoole, Head of the Tumour Bank at the Children's Hospital, Westmead for his support and advice, especially in clinical domain. He provided the project with necessary clinical resources, and orchestrated the clinical team which participated in my PhD.

I would like to express great appreciation to Dr. Sedighe Vajar from Marzdaran pathology laboratory, my co-supervisor Dr. Karla Felix Navarro, Faculty of Engineering and Information Technology, Dr Nicole Graph, Head of the Department of Histopathology Children's Hospital Westmead, and Drs Amanda Charlton and Susan Arbuckle senior specialist anatomical pathologists from the Department of Histopathology, Children's Hospital, Westmead. I would like also to thank Ms. Birgit Smith for helping me to identify and correct grammar, syntax and presentation problems in my thesis.

Most of all I would like to extend my deepest gratitude to my parents Sedighe and Kambiz who made this possible with their constant support. Thank you for always being there for me during the ups and downs and for believing in me more than I do in myself. My past and my future are due to your devotion. I hope there will be a chance to compensate you for that.

Abstract

The goal of this thesis is to develop an innovative Computer Aided Diagnosis (CAD) system for the common deadly infant cancer of Neuroblastoma. Neuroblastoma accounts for more than 15% of childhood cancer deaths, and it has the lowest survival rate among the paediatric cancers in Australia. In quantitative analysis the total number of different regions of interest are counted, and qualitative analysis determines abnormalities within the tumour.

Quantitative and qualitative analysis of tumor samples under the microscope is one of the key markers used by pathologists to determine the aggressiveness of the cancer, and consequently its therapy. Because of the variety of the histological region types and histological structures in the tissue, analyzing them under the microscope is a tedious and error-prone task for pathologists. The negative effects of inaccurate quantitative and qualitative analysis have led to an urgent call from pathologists for accurate, consistent and automated approaches.

Computer Aided Diagnosis (CAD) is an automated cancer diagnostic and prognostic system which enhances the ability of pathologists in the quantitative and qualitative analysis of tumor tissues. However, there are four main issues with developing a CAD system for pathology labs: First is the fluctuating quality of the histological images. Second is a wide range of different types of histological regions and histological structures with complex morphology each adopting a

specific algorithm. Third is overlapping cells which decrease the accuracy of quantitative analysis. Fourth is a lack of utility for pathology labs when they do not follow an appropriate clinical prognosis scheme. Moreover, most of the proposed CAD systems perform either quantitative or quantitative analysis and only very few of them manipulate both types of analysis on the cancerous tumor tissue.

This thesis aims to address the issues raised by developing an innovative CAD system that assists pathologists in determining a more appropriate prognosis for the leading infant cancer of Neuroblastoma. The CAD will automatically perform quantitative and qualitative analysis on images of tumor tissue to extract specific histological regions and histological structures which are used for determining the prognosis for Neuroblastoma.

This thesis has four main contributions. Contribution 1 develops novel algorithms to enhance the quality of histological images by reducing the wide range of intensity variations. Contribution 2 proposes a series of segmentation algorithms for extracting different types of histological regions and histological structures. Contribution 3 addresses the issue of overlapping cells by developing algorithms for splitting them into single cells. Contribution 4 grades the aggressiveness level of neuroblastoma tumor by developing a prognosis decision engine.

The main outcomes of the proposed CAD system in this thesis are a series of novel algorithms for enhancing the quality of the histological images and for segmenting histological regions and histological structures of interests, introducing a prognosis decision engine for grading a neuroblastoma tumor based on a well established histopathological scheme, facilitating the process of prognosis and tumor classification by performing accurate and consistent quantitative and qualitative tissue analysis, and enhancing digital pathology by incorporating a

digital and automated system in the work flow of pathologists.

The performance of all the developed algorithms in this thesis in terms of correctly extracting histological regions, histological structures and grading the level of tumor aggressiveness, is evaluated by a pathologist from the department of histopathology in the Children's Hospital at Westmead, Sydney. Moreover, all the results are compared with state of the art methods. The results indicate that the algorithms proposed in this thesis outperform state of the art quantitative and qualitative methods of analysis.

Publications Related to this Thesis

Below is the list of the peer-reviewed journal and conference papers associated with my PhD research:

Refereed International Journals

1. **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole, and P. J. Kennedy. "Non-parametric and integrated framework for segmenting and counting neuroblastic cells within neuroblastoma tumor images." *Medical and Biological Engineering and Computing*, Springer, 51:645–655, 2013. (PMID: 23359256). This paper specifically addresses Chapter 4 and Chapter 5.

2. November 2013: submitted to journal of IEEE Transactions on Information Technology in Biomedicine, **S. Tafavogh**, D. R. Catchpoole and P. J. Kennedy, "Cellular quantitative analysis and splitting overlapping cells using shortest path between critical points on the concave regions". This paper specifically addresses Chapter 5.

3. December 2013: submitted to journal of BMC Bioinformatics, **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy "Segmentation and quantitative analysis of ganglion cells using histological and topological filters".

4. January 2014: submitted to journal of Microarray, A. Chetcuti, N. Mackie, **S. Tafavogh**, N. Graph, A. Charlton, T. Hinwood, D. R. Catchpoole, "Can archival tissue reveal answers to modern research questions?: Computer-aided histological assessment of neuroblastoma tumours collected over 60 years". This paper addresses Chapter 6.

Refereed International Conference Publications

1. **S. Tafavogh**, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy (2013). "Segmenting cellular regions of neuroblastoma tumor and splitting overlapping cells using shortest path between convex regions of cell contours", in *A*rtificial Intelligence in Medicine, Springer, LNCS 7885, N. Peek, R. M. Morales and M. Peleg, Eds, Springer. 171-175. This paper addresses Chapter 5.

2. **S. Tafavogh**, D.R. Catchpoole, P.J. Kennedy (2012). "Determining cellularity status of tumors based on histopathology using hybrid image segmentation." IEEE World Congress on Computational Intelligence, Australia, pp:1-8. This paper addresses Chapter 4

3. **S. Tafavogh**. Comparing clustering methods and color-spaces for histological images, in Proceedings of the IEEE International Symposium on Image and Signal Processing and Analysis, pp: 154-59, 2011. This paper addresses Chapter 4.

4. **S. Tafavogh**. "Effect of HSV and L*a*b* color spaces on segmenting histological images by expectation maximisation algorithm", Conference on Computational Molecular Biology and Bioinformatics, Moscow, vol. 22, pp: 359-361, 2011. This paper addresses Chapter 4.

5. **S. Tafavogh**. "The effects of adaptive histogram equalization and color spaces on histological segmentation quality of expectation maximization method". International Conference on Graphic and Image Processing (ICGIP), October 2011, Cairo, Egypt. This paper addresses Chapter 4.

Contents

Table of Symbols

Symbols	Description
n	The total number of pixels in a image
X_i	A pixel i in a image where $i = \{1, ..., n\}$
K	Radially symmetric kernel
h	Mean-shift windows radius
$C_{k,d}$	Positive normalization constant
$\hat{f}_{h_S,h_I}(X)$	Gradient density estimator
G	Kernel window
k and g	Profiles of kernel K and G respectively
m	Mean-shift
X^s	Spatial domain of a pixel
X^i	Intensity domain of a pixel
W_w	Center of kernel w
$X^s_{w,r}$	Spatial domain of convergence points
$Y^i_{w,r}$	Intensity domain of convergence points
V_i	Luminance of pixel i
V	Luminance of white pixel
u'_i and v'_i	Chromaticity coordinates of pixel i
u' and v'	Chromaticity coordinates of white pixel

Symbols	Description
h_s	Spatial bandwidth
h_i	Intensity bandwidth
X^I	Intensity domain of pixel in 3-d vector
X^S	Spatial domain of pixel in 3-d vector
y_r	Mean-shift convergence
B	Total number of constructed mosaics in an image
$\mathbf{M}_a$	Mosaic a of size $d \times 3$ where $a = \{1, ..., B\}$ and d is the total number of pixels within the mosaic.
θ	A set of all mosaics in an image
$\bar{\mathbf{M}}_a$	The unified intensity mosaic
c	The total number of manually labeled pixels in cellular regions
t	The total number of manually labeled pixels in neuropil regions
$\mathbf{C}$	A matrix contains the manually cellular labeled pixels
$\mathbf{N}$	A matrix contains the manually neuropil labeled pixels
Z and L	The total number of pixels within the manually segmented cellular and neuropil regions respectively
$\bar{\mathbf{C}}$	Cellular filter
$\bar{\mathbf{N}}$	Neuropil filter
D_a	Euclidean distance between mosaic a and the filters
P_a	Vector which indicates cellular mosaics
AG_C	Absorption grade for cellular filter
AG_N	Absorption grade for neuropil filter
ρ	A vector of $B \times 1$ which indicates all the mosaics of neuropil regions

Symbols	Description
β	A column vector of length B which indicates all the mosaics of new nuclei regions.
ϑ	Vector which indicates the constituent mosaics of cellular regions (unrefined)
ζ	Vector which indicates the constituent mosaics of neuropil regions within ϑ
ϑ'	Vector which indicates the constituent mosaics of cellular regions (refined)
ϱ	Vector of size B which indicates cytoplasm regions
S	The total number of segmented nuclei
b	The total number of pixels in the centroid of the nuclei
$\mathbf{W}_j$	Centroid nuclei matrix for nuclei image
$\mathbf{Q}_j$	Centroid nuclei matrix for cellular image
$\mathbf{\Delta}_j$	A matrix that contains cytoplasm regions
$\mathbf{av}_j$	The average distance between centroid of nuclei j to background
$\mathbf{\bar{av}}_j$	The average distance of all b pixels of centroid j
$\mathbf{aw}_j$	The average distance of eight neighbor pixels for every pixel at centroid j to background in $\mathbf{W}_j$
$\mathbf{aq}_j$	The average distance of the eight neighbor pixels to background in $\mathbf{Q}_j$ for every pixel at centroid j
$\mathbf{\bar{aw}}$	The average distance of all b pixels of centroid j in **aw**
$\mathbf{\bar{aq}}$	The average distance of all b pixels of centroid j in **aq**
O	The total number of objects in an image
o	An object number

Symbols	**Description**
A_o	Area of object o
DE_o	The ratio of the minor diameter to the major diameter of object o
$\hat{A}_o$	Convex hull of object o
PO	Potentially Overlapping
PS	Potentially Single
SC	Single Cell
OC	Overlapping Cell
F	The total number of overlapping cells
l	Overlapping cells l where $l \in \{1, ..., F\}$
R_l	Entire region of overlapping cell l
$\hat{R}_l$	Convex hull region of overlapping cell l
H	The total number of splitting triangles for overlapping cell l
ST_x	Splitting Triangle $\forall x \in \{1, ..., H\}$
ϕ_l	A set of splitting triangles for overlapping cell l
$e_l(ST_x)$	The edge of splitting triangle x for overlapping cell l
$e_l(\hat{R})$	The edge of convex hull for overlapping cell l
$chord_l(ST_x)$	The edge of chord of splitting triangle x for overlapping cell l
$arc_l(ST_x)$	The edge of arc of splitting triangle x for overlapping cell l
EA	A matrix that contains pixels at the arc of overlapping cell l
EC	A matrix that contains pixels at the chord of overlapping cell l
DA	Distance matrix based on **EA**
DC	Distance matrix based on **EC**
DM	Distance between the pixels at the arc
	and chord of a splitting triangle
SP	The total number of initial splitting points for overlapping cell l

Symbols	Description
$\mathbf{DM}$	Distance matrix
$\mathbf{\Theta}_p$	The distance between object p from the $SP-1$ other objects in an overlapping cell l
$e(R)_l$	The edge of overlapping cell l
$\mathbf{F}_p$	A set of $SP-1$ matrices that have the smallest elements of the matrices in $\mathbf{\Theta}_p$
PS	The point sets
STL	The structuring elements
$\mathbf{T}_p$	The splitting routes for initial splitting point p
$\mathbf{J}_p^*$	A splitting route for initial splitting point p
np	The total number of pixels within the segmented neuropil regions
cp	The total number of pixels within the segmented cellular regions
NR	The amount of neuropil regions within an image
$ganglion$	The total number of identified ganglion cells in the segmented cellular regions
$neuroblast$	The total number of identified neuroblast cells in the segmented cellular regions
GR	The amount of ganglion cells within an image

Chapter 1

Introduction

The aim of this thesis is to develop a series of novel algorithms for a Computer Aided Diagnosis (CAD) system that assists pathologists to determine more accurate prognosis for the leading infant cancer, Neuroblastoma. The proposed algorithms in this thesis automatically perform quantitative and qualitative analysis of specific histological regions and histological structures within the histological images of the tumor tissues.

Neuroblastoma Tumor (NT) is an aggressive, infant cancer of nerve cells which accounts for more than 15% of infant cancer deaths worldwide (Park et al., 2010). The lowest survival rate amongst all other infant cancers in Australia is the result of NTs (Baade et al., 2010). Survival after a diagnosis of cancer is highly dependent on the accuracy of the prognosis. However, for NTs, there is a 20% discrepancy between central pathologists who propose prognosis schemes and the prognosis made by institutional pathologists who use those proposed prognosis schemes (Teot et al., 2007).

As a result, there is an urgent demand for transforming conventional pathology to digital pathology as a new framework for improving patient outcomes

by incorporating computer technology in the work flow of pathologists. Computer Aided Diagnosis, a principal component of digital pathology, improves the accuracy of diagnosis/prognosis and reduces medical error. Computer Aided Diagnosis systems are becoming part of routine clinical work, and they are a new class of automated cancer diagnostic/prognostic systems which have evolved following the recent advent of digital slide scanners, which scan histological glass slides and convert them to digital images.

Automated quantitative analysis of histological slides helps biologists in counting different histological structures of interest. Accurate quantitative analysis reduces the number of medical errors and allows biologists to make more directed diagnosis, prognosis and treatment (Gurcan et al., 2009). Quantitative analysis of digital tissue slides requires the detection and counting of certain histological structures such as cells, nuclei, cytoplasm and glands.

Automated qualitative analysis assists pathologists in determining treatment, diagnosis and prognosis of cancers. In qualitative analysis, the results of quantitative analysis, and presence, extent, combination, ratio and relation between histological regions and histological structures are used as key markers for classifying the malignancy and level of aggressiveness of a cancerous tumor.

Conventional pathology is not well aligned with advances in technology which provide new knowledge about the biology of tumors and diagnosis/prognosis protocols and it is often effected by a degree of inter-intra observer variability in cancer grade determination (Teot et al., 2007). Intra observer variability is differences between pathologists in determining diagnosis, prognosis and treatment for a cancerous tumor wihtin the same institute, while inter observer variability is discordance between diagnosis, prognosis and treatment made by a pathologist and the defined prognosis scheme for a cancer. Analyzing many different types

of histological regions, abnormalities and histological structures under an optical microscope is a tedious and time-intensive task, and is subject to inter-intra observer variability.

The prognosis of a tumor determines the chances of survival for the patient. To develop an automated system for making a prognosis for cancerous tumors, different medical image processing tools and computerized classification rules must be constructed to perform quantitative and qualitative analysis of tumors.

In this thesis, a series of novel algorithms and computerized rules are proposed for automated quantitative and qualitative analysis of the histological images of the NTs. The histological regions which are considered by the proposed algorithms in this thesis, are cellular regions and connective tissues such as neuropil regions. The histological structures are nuclei, cytoplasm, and specific types of nerve cells such as ganglion cells and neuroblast cells. Moreover, the proposed algorithms grades NTs as undifferentiated, poorly differentiated and differentiated based on results of the analysis. Differentiated tumors are less aggressive, while undifferentiated tumors are highly aggressive.

1.1 Conventional Pathology, Associated Errors and Digital Pathology

Pathology is one of the last areas in medicine that is still dependent on analogue devices such as the optical microscope and tissue glass slides. There is considerable debate in the US and Europe about enhancing the techniques of conventional pathology to deal with recent sophisticated knowledge about the biology of tumors. This debate has increased following the release of research outcomes

by Khoury et al. (1996), Troxel (2004) and Valenstein and Sirota (2004) which demonstrate that there is discordance between the defined schemes for making prognosis, and the prognosis which is made by pathologists. Khoury et al. (1996) reported two types of errors in pathology labs; 1) procedural errors, and 2) cognitive errors. Procedural errors include technical failure, equipment problems, and organizational flaws, which are outside the domain of this project. Cognitive errors by pathologists include faulty detection, misidentification of symptoms or signs, synthesis, and perception. This thesis addresses these cognitive errors by; improving the accuracy of detecting different types of histological regions and histological structures, discriminating between different types of cells, and following a histopathological scheme to make an accurate prognosis for NTs.

Significant discussion has taken place in the US and Europe (Cornish et al., 2012; Khoury et al., 1996) about the transformation of conventional pathology to digital pathology as an efficient framework. Advanced hardware and software digital tools facilitate the process of diagnosis and prognosis, and consequently, they enhance both therapy and patient outcomes. Computer Aided Diagnosis is a principle component of digital pathology, and they becomes part of routine clinical work (Gurcan et al., 2009).

1.2 How does a CAD System Analyze Tissue Samples?

The first step in the computerized analysis of tissue specimens is to interpret the histological regions and the histological structures in the tissue by converting them into digital format. With the advent of whole slide digital scanners, slides

can now be recorded in digital image format. The digital images are in numeric form and are constructed by millions of color pixels. Each pixel is defined by a three dimensional function $f(x, y, z)$, where x and y are spatial coordinates of the pixel, and z shows the range of main colors, Red, Green, Blue (RGB) of the pixel at point x, y. Each histological structure has a specific morphology in the real tissue slide, and each has a correspondingly specific numerical pattern in the image.

Medical image processing uses or improves the tools provided by digital image processing and data mining techniques to mine the specific numerical patterns and determine the biological data within the digital images. Segmentation is the main tool used for the separation of structures of interest from the background and from one another (Gonzalez et al., 2004). The segmentation process groups the pixels and identified patterns within the image, allowing the computer to determine abnormalities and specific histological structures based on the numeric data provided by their constituent pixels. The outcomes of the segmentation can be used for quantitative and qualitative analysis.

To determine a prognosis for a cancerous tumor, the CAD performs a qualitative analysis on the tumors. To perform a qualitative analysis, all the segmented histological regions, histological structures and abnormalities must be taken into consideration. Then, based on the prognosis scheme of the tumor a set of rules is defined to classify the aggressiveness of the tumor.

1.3 Research Challenges

This thesis addresses four main challenges for developing CADs that analyze digital histological slides:

Challenge 1 The fluctuating quality of the slides. To allow observation of the different types of histological structures, the tissue on the slides must be stained. There are many methods of staining slides, such as Hematoxylin & Eosin (H&E), Periodic Acid-Schiff stain and Feulgen's stain (Fox, 2000). Each of these staining methods are subject to technical variations that influence the quality of the final stain and the ability to visualize histological regions and histological structures. Hematoxylin & Eosin is the most widely used method because of its diagnostic capability which is enhanced by the strong contrast between structures of different composition, and strongly dissimilar colors of cell nuclei and cytoplasm (Fox, 2000). The main drawback of H&E for CAD systems in histopathology is the fluctuating quality of the staining intensity. For example, the same histological regions, such as cellular regions in different locations of the slide, are stained with different shades of blue. Such color and intensity variation reduces the performance of the CAD in detecting cellular structures.

Challenge 2 Locating and extracting different histological regions and histological structures. Identifying different types of histological regions and histological structures within the digital image according to their specific numerical pattern, is the most complicated stage of the automated quantitative and qualitative analysis. Usually, each individual type of histological structure has its own numerical pattern and requires an independent image processing approach. Neuroblastoma tumor contains different histological regions such as cellular regions, connective tissues, neuropil regions, red blood cells and background. Each of them requires specific hybrid algorithms for segmentation purposes.

Challenge 3 Addressing the negative effects of overlapping cells on the performance of cellular quantitative analysis. A fixed biopsy on a glass slide is a 2-D section of 3-D tissue. Transforming tissue samples from 3-D to 2-D space results in overlapping cells. In medical image processing, each cell is considered to be an object, and because overlapping cells are composed of two or more connected cells, they are considered as a single object. A tissue image with an enormous number of overlapping cells, significantly reduces the accuracy of the segmentation (Gurcan et al., 2009).

Challenge 4 Inaccurate qualitative analysis of the tumors which increases the medical errors and lack of CAD utility for clinical laboratories. The main reasons are these systems do not allow segmentation of all the necessary histological regions and histological structures, and they do not accurately follow a clinical prognosis scheme.

1.4 Research Aims and Contributions

By analyzing the main challenges of the given domain, this thesis identifies four main contributions:

Contribution 1: Global and local intensity variation reduction approach to enhance the quality of histological images. The fluctuating quality of H&E staining has a negative effect on segmentation performance, as it provides different numerical patterns for the same type of histological structures in the images. In these cases, similar objects have similar numerical patterns and different numerical patterns represent different objects. Thus, the system mistakenly classifies two identical histological structures as two

different histological structures. Once the image is loaded to the system, its quality needed to be enhanced by reducing the color variation in the image thereby increasing the readability of the image for the system. In Chapter 4, this issue is addressed by proposal of robust global and local intensity variation reduction algorithms. The chapter explains the proposed approaches for reducing the intensity variation through the entire images. It also explains the approaches for reducing the number of distinct colors within the image by clustering the pixels according to their intensity and spatial domain similarities. Lastly, colors of the pixels in each of the clusters are unified to the mean value of the colors of all pixels within that cluster.

Contribution 2: Novel approaches to segmenting histological regions of interest. Usually, each type of histological region and histological structure has its own numerical pattern and requires a separate image processing approach. In this thesis, different algorithms based on novel techniques are proposed to locate and extract the cellular regions, connective tissues such as neuropil regions, neuroblast cells, ganglion cells, nuclei and cytoplasm structures in NT. Chapter 4 explains the algorithms for segmenting these histological regions and histological structures. The accuracy of the algorithms in terms of segmenting the histological regions and histological structures is evaluated by a pathologist. Moreover, the performance of the system in terms of segmentation is compared with other robust state of the art methods.

Contribution 3: Identifying and splitting overlapping cells To discriminate

overlapping cells from single cells, and split the overlapping cells into single

cells, a two stage algorithm is developed by: 1) Analyzing morphological differences such as area, minor and major diameters, and concavity of the cells between overlapping cells and single cells to distinguish between overlapping cells and single cells. 2) Identifying the initial splitting points by deploying the introduced concave region determination, decomposition, seed growing and critical points acquiring techniques. 3) Determining a shortest path method to identify a split route between the candidate pair of initial splitting points. In Chapter 5, the details of the algorithms are explained. The accuracy of the algorithms in terms of determining overlapping cells and splitting them into single cells is evaluated by a pathologist. Also, the performance of the algorithms in terms of splitting the overlapping cells and counting the single cells is compared with other state of the art methods.

Contribution 4: A prognosis decision making engine to classify the aggressiveness of NT. To grade the aggressiveness of this cancerous tumor, a prognosis decision engine is proposed to compare the relationship between the segmented histological regions and histological structures according to a clinically established prognosis scheme. The prognosis decision engine contains a set of computerized rules based on procedures and considerations used by pathologists to determine the prognosis and diagnosis for the NTs. The accuracy and utility of the proposed prognosis decision engine is evaluated by a paediatric pathologist. Moreover, the performance of the system is compared with another proposed CAD system. The results and statistical tests indicate a reliable and superior performance in grading the aggressiveness of the NTs by the proposed CAD in this thesis. Moreover, statistical tests show that the proposed CAD in this thesis significantly out-

performs other CADs in the state of the art in determining the grade of NTs.

1.5 Research Significance

Neuroblastoma tumor is the most aggressive infant cancer with the lowest survival rate among paediatric cancers, and the three-year survival rate for this cancer is less than 30% around the world (Baade et al., 2010). Thus, improving the process of diagnosis, prognosis and treatment for NTs, enhances patients chance of survival. Moreover, well defined morphological features, a grading scheme and a manageable number of samples of NTs allow us to design and develop precise algorithms and computerized classification rules based on a robust pattern.

This thesis proposes a new platform for quantitative and qualitative CAD systems. The proposed CAD system improves digital pathology by performing new series of quantitative and qualitative analysis on the tissue images. The CAD system provides an accurate prognosis tool which improves the confidence and accuracy of pathologists in determining the treatment for the cancer. The system facilitates the process of prognosis for pathologists. The CAD has the potential to reduce the number of hours pathologists spend on quantitative and qualitative analysis of cancerous tumors, allowing them to work on more sophisticated cases. Finally, the CAD system improves early detection and diagnosis of one of the most aggressive types of infant cancer. By applying some modifications, the proposed algorithms and approaches in this thesis are scalable to other CAD system for locating and segmenting other histological regions and histological structures.

1.6 Research Hypothesis

The research hypothesis is to provide a robust and accurate CAD system which performs quantitative and qualitative analysis on the NTs. The system assists pathologists to decrease the discordance between the existing prognosis schemes and the prognosis made by pathologists. The experiments results show that the developed CAD in this thesis provides the best results in terms of quantitative and qualitative analysis of NTs in comparison with state of the art systems.

1.7 Thesis Structure

The thesis contains seven chapters as follows:

- Chapter 1 presents the research background, research challenges, research objectives and contributions, research significance and the overall research methodology.

- Chapter 2 indicates the literature relevant to this study. It includes a review on biological domain of NTs, algorithms and automated systems for quantitative and qualitative analysis, and gaps with current.

- Chapter 3 shows the overall model of the proposed CAD system.

- Chapter 4 develops algorithms for the pre-segmentation process, to enhance performance in locating and extracting the histological regions and histological structures of NT, and consequently improving the performance of the system in quantitative and qualitative analysis. This chapter introduces a series of algorithms for locating histological regions such as neuropil and cellular regions, and segmenting histological structures such as nuclei, cyto-

plasm, neuroblast cells and ganglion cells. To validate the performance of the proposed algorithms, their results are evaluated by a pathologist and all the algorithms are compared with the associated state of the art algorithms.

- Chapter 5 describes the proposed novel algorithms for the post-segmentation process to improve the quantitative analysis of neuroblast cells. The results of the proposed algorithms are evaluated by a pathologist and their performance are compared against the performance of state of the art algorithms.

- Chapter 6 classifies the aggressiveness level of NTs by developing an automated prognosis decision making engine which performs the qualitative analysis. The prognosis results of the system are evaluated by a pathologist. When the results are compared with those of the other CAD system for the prognosis of NTs, the superiority of the proposed system is indicated.

- Chapter 7 provides conclusions, discussions and future research directions for the work presented in this thesis.

Chapter 2

Literature

This chapter reviews literature related to this project. Section 2.1 explains the biological domain of NTs, the histological regions and histological structures within the tumor, the grading scheme of NTs and medical error related to the prognosis of NTs. Section 2.2 reviews developed algorithms and systems developed for automated quantitative and qualitative analysis. Finally, Section 2.3 explains the research gaps for developing a CAD system for NTs.

2.1 Biological Domain

Neuroblastoma tumors are embryonal malignancies of the sympathetic nervous system which originate from the neural crest. The prognosis of NT is related to several factors: 1) age of the patient, 2) location of the tumour, 3) surgical staging, and 4) microscopic grading (Rosai et al., 2004). The first three factors can usually be determined accurately and easily by pathologists; however, determining microscopic grading requires extensive analysis of different histological regions and histological structures within tumor tissue under the microscope.

2.1.1 Histological Regions and Histological Structures of NT

Some of the important histological regions that are considered by pathologists in making a prognosis for the tumor are as follows:

- Cellular regions: in the H&E stained tissue slide, cellular regions are stained blue as shown in Figure 2.1.
- Neuropil regions: in the H&E stained tissue slide, neuropil regions are stained pink as shown in Figure 2.1. Neuropil is a region between neuronal cell bodies in the gray matter of the brain and spinal cord. Neuropil is an aggregation of cell membrane processes from surrounding neuronal cells. Less aggressive types of NTs contain discernible amount of neuropil.

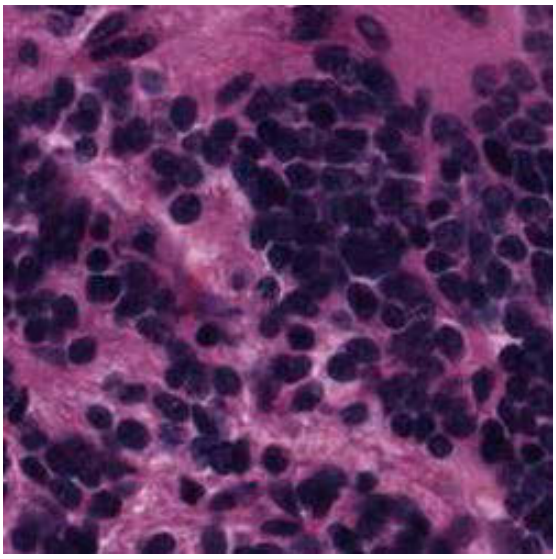


Figure 2.1: Blue regions are cellular regions and pink regions are neuropil regions in a H&E stained histological image.

The histological structures within the histological regions which are considered by pathologists for making a prognosis for NTs are as follows:

- Neuroblast cell: small to medium in size, with indiscernible to thin cytoplasm and vaguely defined cytoplasmic borders. The dark blue circles in Figure 2.2a are neuroblast cells.

- Ganglion cell: has an enlarged, eccentric nucleus with a single prominent nucleolus and conspicuous cytoplasm. The size of this cell is more than twice the size of its nucleus as shown in Figure 2.2b.

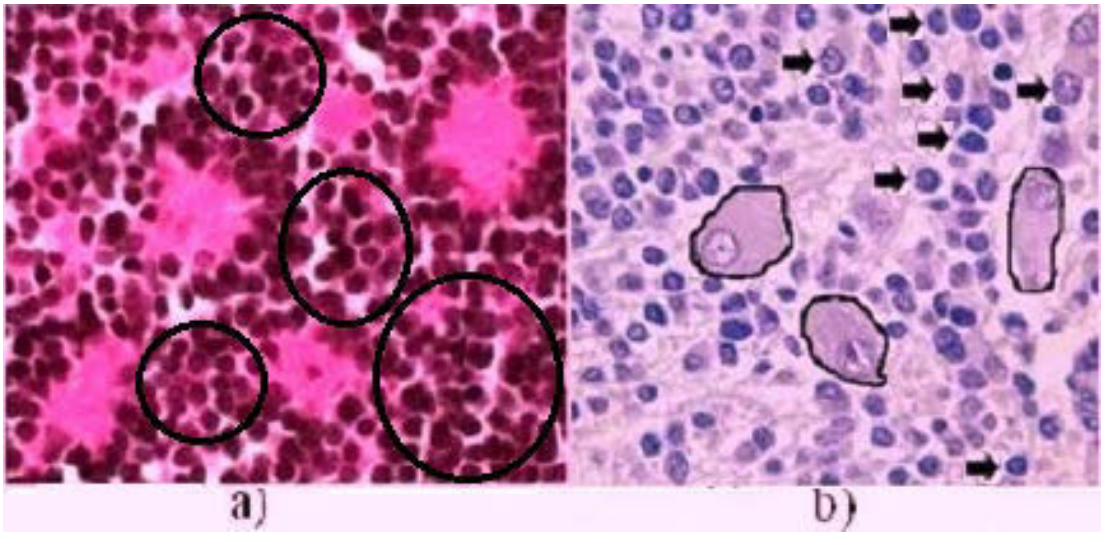


Figure 2.2: a) The dark blue spots inside the circles are neuroblast cells in a H&E stained image, b) lines indicate ganglion cells in a H&E stained image, and arrows point to neuroblast cells.

2.1.2 Prognosis Scheme

Each tumor has a certain prognosis scheme which allows pathologists to determine the aggressiveness level of the tumor and the chance of survival for patients. A prognosis scheme has a set of rules for pathologists to determine tumor aggressiveness based on different combinations of histological regions and histological structures. Over the years, several prognosis schemes have been proposed for NTs, largely based on microscopic markers such as histological regions and histologi-

cal structures. For example prognosis schemes have been proposed by Beckwith and Martin (1968), Hughes et al. (1974), Joshi et al. (1992), and Shimada et al. (1999). Amongst these prognosis schemes, Shimada et al. is the newest, most comprehensive, and widely used scheme (Rosai et al., 2004). According to Shimada et al., NTs are classified into four categories and their subtypes, so helping pathologists to determine aggressiveness of the tumor:

1. Neuroblastoma with undifferentiated, poorly differentiated, and differentiated subtypes.

2. Ganglioneuroblastoma intermixed.

3. Ganglioneuroma maturing and mature.

4. Ganglioneuroblastoma nodular.

Figure 2.3 indicates different grades of NTs based on Shimada classification.

This thesis addresses the first type of NTs which is the Neuroblast type. To determine other types, a combination of both microscopic and macroscopic analysis is required. Macroscopic analysis is beyond the capabilities of the CAD systems. According to Shimada et al. (1999) to determine the subtypes of the first category (Neuroblastoma), pathologists consider the combinations of the following histological regions and histological structures for each of the subtypes:

1. The **undifferentiated** subtype contains cellular regions which have an histological structure of mainly neuroblast cells. An identifiable background of neuropil is absent. Undifferentiated subtype has low patient chance of survival.

2. The **poorly differentiated** subtype is defined as a tumor with a background of more than 2% neuropil regions, and its presence is readily apparent in a tumor tissue sample. Moreover, 5% or less of the tumor cellular regions has

an histological structure of ganglion cells. Poorly differentiated subtype has low patient chance of survival.

3. The **differentiated** subtype has abundant neuropil and 5% or more of the tumor cellular regions showing ganglion cells. This subtype has good patient chance of survival.

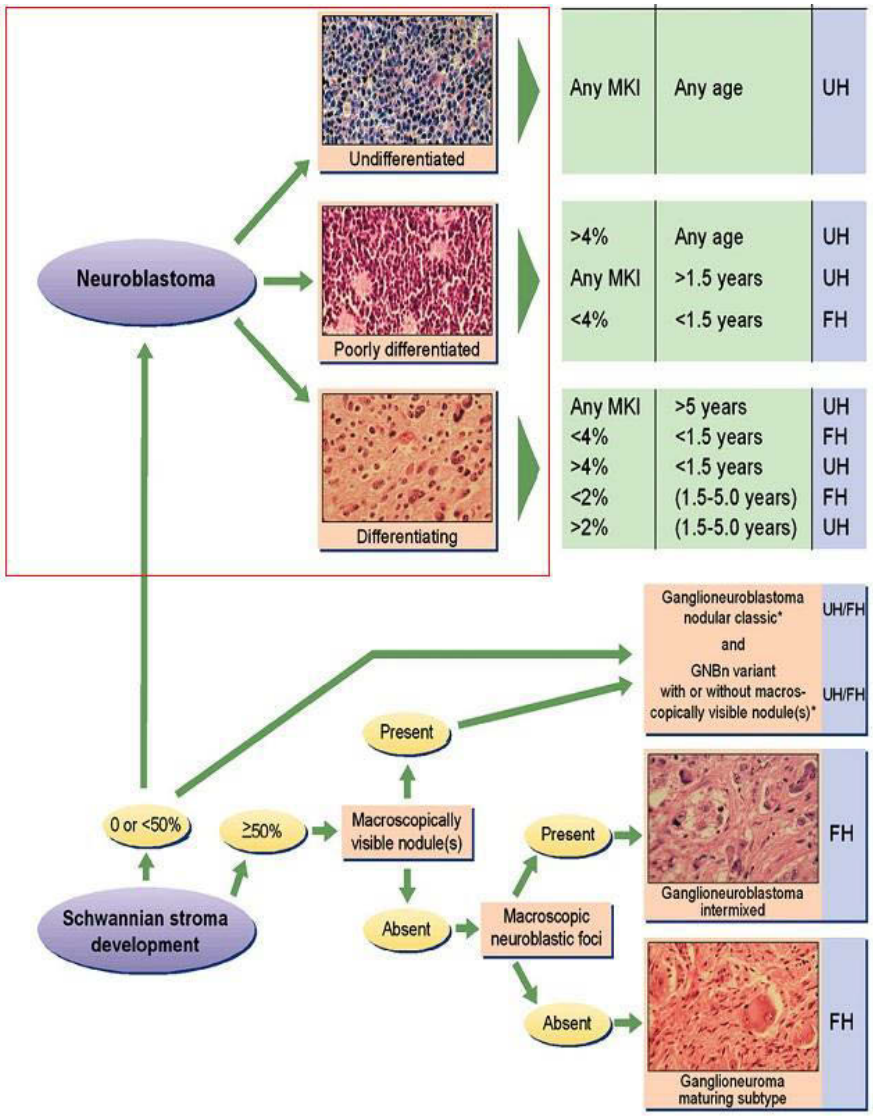


Figure 2.3: Different grades of NT (Park et al., 2010). The red rectangle indicates the part of the grading scheme which is covered by this thesis. MKI indicates the ratio of Mitosis cells to Karyorrehectic cells, UH is Unfavorable Histology and FH is Favorable Histology.

2.1.3 Errors in Pathology

There is considerable debate about the accuracy of pathologists in determining the microscopic grading of NTs. Teot et al. (2007) indicate that there is a 20% cognitive discordance between the prognosis scheme introduced by central pathologists and the prognosis made by institutional pathologists for NTs. Khoury et al. (1996) reported that these cognitive errors are the results of faulty detection, information processing, synthesis and perception. As a result automated systems that accurately and robustly perform quantitative and qualitative analysis on the tumor tissue images are urgently needed by pathologists (Teot et al., 2007). In the following sections, the proposed algorithms, methods and systems to address the above-stated demand are explained.

2.2 Algorithms and Automated Systems

Researchers both in the medical image analysis and pathology fields have recognized the importance of quantitative and qualitative analysis of histology images (Gurcan et al., 2009). Different systems and algorithms have been proposed for this quantitative and qualitative analysis. In this section of the thesis the proposed algorithms and systems for automated quantitative and qualitative analysis and their drawbacks are explained.

2.2.1 Algorithms and Systems for Quantitative Analysis

In this section algorithms and systems which were proposed for locating and segmenting different histological regions and histological structures are discussed. These algorithms and systems perform quantitative analysis on the images of the

tumor tissues. In next sections:

1. Thresholding and watershed algorithms for locating and segmenting different histological structures are reviewed.

2. Algorithms based on morphological analysis are explained.

3. Algorithms and systems based on clustering techniques to identify different histological structures are introduced.

4. Systems based on active contour model are reported.

5. Post-segmentation technique to improve the accuracy of quantitative analysis are discussed.

Thresholding and Watershed Techniques

Thresholding is a well known technique for locating and segmenting histological regions and histological structures within tumors (Gonzalez et al., 2009). The principle idea behind the thresholding technique is to binarize the image into a foreground/background image using a grey level histogram of the pixels. Thresholding techniques provide high performance on images with a uniform intensity of distribution. The main drawback of thresholding is its poor performance in detecting regions of interest where an image has many different regions.

Watershed (Roerdink and Meijster, 2000) is segmentation technique which is widely used for segmenting histological structures. The main drawback of watershed technique is high rate of over-segmentation. Watershed technique follows four simple steps:

1. Finding minima points of the regions using distance transformation (Gonzalez et al., 2009) and assigning a unique identification label to each minima

point. The minima points are called "seeds" in the watershed technique.

2. Growing the seeds in an iterative process and assigning the same identification label to each of the corresponding growing regions.

3. Terminating the process of growing as soon as regions with different identification labels meet.

4. Separating two regions at the meeting point by a watershed line.

Ta et al. (2009) developed a hybrid algorithm which incorporates both the thresholding and watershed techniques, to segment the cell nuclei clusters from peripheral blood and bone marrow preparation. The system applies the thresholding proposed by Kittler and Illingworth (1986) to fluorescence microscopic images after a shading correction and background subtraction have been performed. The principle idea behind the Kittler method is to optimize a criterion function related to the average pixels classification error rate. It chooses a grey level threshold between the dominant intensity values of objects and background intensities to segment images. The system then deploys the watershed technique to disconnect the touching nuclei. The main drawback of the system is the sensitivity of the thresholding technique to wide range of color variation which significantly reduces the performance of the system. Moreover, high over-segmentation reduces the performance of the system in segmenting cell nuclei.

Coelho et al. (2009) implements histogram equalization (Gonzalez et al., 2009) to reduce the color variation as a pre-segmentation operation. This improves the performance of the thresholding in the images with wide range of intensity variation. The system then integrates thresholding and watershed techniques to perform an objective evaluation of the nuclei segmentation algorithms. The outcome

of the system indicates that the application of the pre-segmentation algorithm improves the performance of the thresholding. Moreover, Russell et al. (2009) proposed another method to overcome the low performance of the thresholding technique in the images with wide range of intensity variations. The system uses an adaptive thresholding technique and watershed technique to extract the cellular regions and segment the mammalian cell nucleus. The adaptive technique automatically determines the threshold level for the mammalian cell nucleus. The results indicate that the proposed method significantly improves the performance of the thresholding technique by obtaining closer results to the ground-truth than the traditional thresholding method proposed by Kittler and Illingworth (1986). The proposed adaptive thresholding and watershed technique has also been used by Madhloom et al. (2010) to distinguish five different types of white blood cell nuclei. However, over-segmentation still remains a drawback of the system.

The system proposed by Zhou and Li (2009) integrates thresholding and watershed techniques. It addresses the drawbacks of both techniques by using adaptive thresholding, a novel hybrid merging algorithm, to reduce the negative effects of high over-segmentation derived by the watershed technique. This algorithm combines a compactness score with the probability distribution function score to merge the fragmented cells produced by the watershed technique. The compactness score indicates the shape of the nuclei. It means that a nuclei with more round shape has smaller compactness score while a fragmented nuclei has high compactness scores. The results indicate an improvement in performance of the watershed algorithm. However, by comparing the results with those of other segmentation algorithms, such as an algorithm based on morphology, geometry and clustering, the system still suffers from over-segmentation.

This thesis introduces algorithms for segmenting cells with low over-segmentation

drawback. In section 5.4 the performance of the system proposed by Zhou and Li (2009) in segmenting cells is compared with performance of the system proposed in this thesis.

Morphological Analysis and Geometrical Model Fitting

Morphological analysis and geometrical model fitting take different morphological features, and the shape and geometry of the objects into account. Because histological regions do not follow any specific morphology, the above techniques are used to locate and segment the histological structures such as specific types of cells or glands. Some of the morphological and geometrical parameters that are taken into consideration in the techniques are convexity, concavity, size, area, edge, convex-hull, bounding box of the objects, and ellipse fitting. The main challenges with these systems are noise in the morphology of the histological structures (such as fluctuating contour of the cells), and unclear borders between two histological structures (such as overlapping cells).

Yeo et al. (1994) propose a system based on morphological and geometrical analysis to segment a specific type of cells, by calculating the amount of convexity on the contour of the cells. The system provides high performance if the contour of the cells has a uniform distribution, although cells mostly have fluctuating contours.

Kovalev et al. (1996) used shape based analysis to extract the nuclei of cellular regions within the fluorescent stained histological images. This decreases the negative impact of fluctuations of the contours of cells. The system uses thresholding technique to detects cellular regions. To determine the nuclei within the cellular regions it then uses dominant concave regions of the segmented nuclei by a set of geometrical templates. The system then determines the correlation between

the obtained concave regions and the geometrical templates, and considers those concave regions as nuclei if their correlation is above a certain threshold. The main drawback is that the system is unable to detect overlapped nuclei within the image and as a result it considers them as a single nucleus. This significantly reduces the accuracy of the system for quantitative analysis.

Cong and Parvin (2000) proposed a system to extract the nucleic structures within the cellular regions. The system implements adaptive thresholding to extract the cellular regions. It then uses polygonal fit to extract the nuclei structures from the cellular regions. To deal with the overlapping nuclei, the system employs the concavity of each fitted polygon to split the touching nuclei. Finally the system uses the hyperquadric model to determine the boundaries of all the split nuclei within the image. The key issues with this method are low performance of the system in splitting the clump of nuclei with more than two overlapped nuclei, and the complex computations. This is because the system confuses the critical points on the arches and chords of convex regions, and consequently cannot correctly identify the splitting path.

Bai et al. (2009) proposed a system to deal with locating and segmenting nuclei from clump of nuclei with more that two touching nuclei. The proposed algorithm for the system contains two parts, contour pre-processing and ellipse processing. For contour pre-processing part the system first applies an opening morphological operation (Soille, 2003) followed by a closing morphological operation (Soille, 2003) to smooth the contour of nuclei. The system then uses the adaptive thresholding technique based on the Otsu method (Otsu, 1975) to form a binary image. Lastly, the system extracts the contour of nuclei using the watershed technique. In the ellipse processing part, the main purpose of the system is to split the touching nuclei into different ellipses which represent separated

nuclei. After the contour has been segmented, ellipse fitting is used to fit the contour. Therefore, overlapping nuclei are separated by isolated ellipses. Although the application of opening and closing operations reduces the over-segmentation which is a drawback of watershed technique, this system still has minor over-segmentation. The watershed technique fragments cells into a number of pieces, and each of these pieces is considered by the ellipse fitting operation to be an ellipse and consequently a nucleus. Therefore, the system counts more nuclei than the original number of nuclei in the image.

Lu et al. (2013) proposes an algorithm in three steps. The first step involves extracting cellular regions using maximally stable extremal regions algorithm (Matas et al., 2004). For the second step, the system incorporates a level set function (Li et al., 2005) with an energy function that is constrained by the individual contour area and length (Li et al., 2010) to detect the nuclei structures. Level set is a non-parametric technique for tracking curves and surfaces on a fixed Cartesian. Finally, to split the overlapping nuclei, an ellipsoidal shape prior (Rousson and Paragios, 2002), and the area and gray values within the overlapping nuclei are used. The main drawback of the system is poor performance in delineating and splitting overlapping nuclei structures with severe overlap and poor contrast.

In Chapters 4 and 5 novel algorithms are proposed to overcome the issues with sensitivity of algorithms to the fluctuating contours of the cells, and overlapping cells. Moreover, performance of the algorithms proposed by Lu et al. (2013) in terms of segmenting cells are compared with the proposed algorithms in this thesis in Chapter 4.

Clustering Techniques

Clustering is a segmentation technique which is very useful for extracting the objects or regions which do not have specific morphological properties such as histological regions. In clustering techniques, images are considered as feature-spaces and the intensity, energy, spatial domain and gradient of the pixels in the image are considered as the features. Then, homogeneous pixels are grouped in the same cluster and different clusters are discriminated by a unique label. Each cluster contains the constituent pixels of an object or part of a region in the image. The main issue with clustering technique is misclassifying histological regions within the images.

Begelman et al. (2004) developed a hybrid segmentation technique to extract the cellular regions and red blood cells from the images by integrating a multi-resolution approach with K-means clustering technique. In the first step, the system adaptively tunes the scale of the image to find the best resolution for enhancing the process of clustering. Then using K-means clustering, the constituent pixels of different histological regions are grouped in different clusters. The system uses Canny edge detection (Canny, 1986) to extract the edges of the cellular regions, and finally it applies a second-order polynomial fitting algorithm to distinguish the nuclei of the cellular regions. The main drawback of the system is its high dependency on user interaction for setting the parameters. The user must define the number of different types of histological structures within the tumor. Moreover, the high standard deviation of the system is an indicator of its fluctuating performance in segmenting different histological regions in different images.

Yang et al. (2006) proposed a cellular region segmentation method to extract the cellular regions and detect the stage of cell mitosis. The system introduces a hybrid algorithm by integrating a classical Gaussian kernel based on mean-shift (Comaniciu and Meer, 2002) with kernels having scale, shape and direction adaptation. Because mean-shift is used in the developed algorithms in this thesis, details of the algorithm are explained in the following. According to Comaniciu and Meer (2002) mean-shift clusters the pixels using their intensities by a multivariate kernel density estimator based on radially symmetric kernel $K(X)$ and window radius h and is computed for a given pixel X by,

$$\hat{f}_{h,K}(X) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{X - X_i}{h}\right) \tag{2.1}$$

where X_i is a pixel in $d-$dimensional space $\mathbb{R}^d$ and n is the total number of pixels in the image. K is satisfied by:

$$K(X) = C_{k,d} k\Big(\big\|X\big\|^2\Big) \tag{2.2}$$

where $k(X)$ is called the profile of the kernel only for $X > 0$ and $C_{k,d}$ is a positive normalization constant that makes $K(X)$ integrate to one. Clusters locate at the mode of the density. Thus, in order to find the clusters the modes must be identified. The modes are located among the zeros of the gradient density estimator ($\bigtriangledown \hat{f}_{h,K}(X) = 0$) which can be calculated using expression (2.1) as follows,

$$\hat{\bigtriangledown} f_{h,K}(X) = \frac{2C_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (X - X_i) g\left[\left\|\frac{X - X_i}{h}\right\|^2\right] =$$

$$\frac{2C_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n} X_i g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)} - X\right] \tag{2.3}$$

where $g(X) = -k'(X)$. Then, the expression (2.3) can be divided into two separate expressions. The first expression is

$$\hat{f}_{h,G}(X) = \frac{C_{g,d}}{nh^d}\left[\sum_{i=1}^{n} g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)\right] \tag{2.4}$$

which it is proportional to the density estimate at X computed with kernel $G(X) = C_{g,d}g(\|X\|^2)$ and the second expression is the mean-shift vector as

$$m_{h,G}(X) = \frac{\sum_{i=1}^{n} X_i g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)} - X \tag{2.5}$$

which indicates the difference between the weighted mean using the kernel G and X as the center of the kernel (window). Then by applying expressions (2.3), (2.4) and (2.5) mean-shift can be rewritten as

$$m_{h,G}(X) = \frac{1}{2}h^2 C\frac{\bigtriangledown f(X)_{h,K}}{\hat{f}(X)_{h,G}} \tag{2.6}$$

The mean-shift vector $m_{h,G}(X)$ leads the window G toward the increase in the density. By successive computation of $m_{h,G}(X)$ the center of the window (X) is shifted toward the stationary point which is the mode of the density. Considering pixel intensity as the only parameter for clustering homogeneous pixels which are distributed in different locations of the image results in misclassification. To

overcome this drawback another parameter relating to spatial domain can be added to the feature space (Comaniciu and Meer, 2002). To do this the radially symmetric kernel $K(X)$ of the density distribution function should be redefined by two radially symmetric kernels as explained in Comaniciu and Meer (2002),

$$K_{h_s,h_i}(X) = \frac{C}{h_s^2 h_i^3} k\left(\|\frac{X^s}{h_s}\|^2\right) k\left(\|\frac{X^i}{h_i}\|^2\right) \tag{2.7}$$

where C is a normalization constant, h_s is the spatial bandwidth, h_i is the intensity bandwidth, X^s is spatial domain of a pixel and X^i is the intensity value of a pixel. By redefining the kernel of mean-shift with expression (2.7) the stationary point (convergence point) will be based on the joint the spatial-intensity domain. Let $W_w = (X^s_{w,r}, Y^i_{w,r})$ be a point that shows the center of the new kernel (window), where $X^s_{w,r}$ indicates the spatial domain and $Y^i_{w,r}$ shows the intensity of the convergence point. Those W_w in the image that are closer than h_s in spatial domain and h_i in the intensity domain are grouped into the same cluster and the pixels inside their window receive the identical label. The pixels with identical labels form a homogenous tile in the image. The system proposed by Yang et al. (2006) is fully automated and provides high accuracy in segmenting cellular regions and nuclei structures. One of the reasons for the high performance is the application of mean-shift as a robust clustering method to group the homogeneous pixels. However, the system has a high standard deviation in analyzing different histological images which means that it is sensitive to noise and has low quality of the images.

Colantonio et al. (2007) introduced a non-parametric clustering algorithm by deploying fuzzy logic to segment different types of histological regions in the H&E stained histological images. The fuzzy rules are based on shape properties, nor-

malized cross correlation with nuclei template and color features, and a mixture of Gaussian distributions. The system is fully automated with a low standard deviation. However, the performance of the system in terms of accurately segmenting different types of histological regions is low.

Ihlow et al. (2011) proposed a system for segmenting the cellular regions of cervical cancer. The system used the Expectation Maximization (EM) clustering method (Dempster et al., 1977) to distinguish between different types of histological regions. The system provides high accuracy in terms of discriminating and segmenting cellular regions. However, the main drawback of the system is the misidentification of those pixels in histological regions which have close ranges of intensity levels to the constituent pixels of other types of histological regions. This results in misclassifying other types of histological regions as cellular region, or in misclassifying cellular regions as other types of histological regions.

Tsai et al. (2013) proposed a method based on integrating fuzzy c-means clustering (Nock and Nielsen, 2006) and the level set method, to segment colon cells. To predict the number of classes for fuzzy c-means this framework applies the Bayes classifier (Devroye, 1996) based on mixture of density models to the constituent pixels of images. Each of the classes indicates a type of histological regions. Then, for segmenting the cellular regions a supervised kernel based fuzzy c-means algorithm using the obtained prior knowledge of the number of classes is applied to images. The system is independent of user interaction for setting the parameters, and the reported results show a high performance for the system in terms of segmenting cellular regions. However, the accuracy of the Bayes classifier in determining the number of classes plays a crucial role on the overall performance of the system because each class belongs to one type of histological region and determining the wrong number of classes means detecting

the wrong types of histological regions. Images with wide ranges of intensity values reduce the accuracy of the Bayes classifier and consequently reduce the performance of the system in segmenting cellular regions. Moreover, the system is computationally complex.

In Chapter 4 mean-shift clustering technique is used to group the homogeneous pixels and a novel pre-segmentation technique is proposed to deal with the issue of misclassifying. Moreover, in Section 4.2.4 a modification to the algorithm of Ihlow et al. (2011) is proposed to segment the neuropil regions instead of cellular regions. In Ihlow et al. (2011) EM classifies all histological regions, thus the algorithms are modified in a way that they show the classified neuropil regions only. Then, the performance of the algorithms proposed by Ihlow et al. (2011) in terms of segmenting neuropil regions is compared with the proposed algorithms in this thesis to show the differences between the two algorithms in classifying histological regions. The results indicate that the proposed algorithms in this thesis significantly overcome the issue of misclassifying the histological regions within the H&E images.

Active Contour

The active contour method (Kass et al., 1988) is a robust technique for locating and segmenting histological structures within the histological regions. Active contour models are energy-minimizing curves which deform to fit image features. The curves lock on to local minima at the edges of two different objects which have two different level of energy. Moreover, a smoothing process keeps the model continuous and smooth, and prevents the creation of fluctuated contours and sharp corners. Most of the algorithms proposed in active contour model deal with two main challenges, 1) objects with overlapping contours, and 2) density

inhomogeneous images. Some of the algorithms for dealing with these issues are reviewed below:

Bamford and Lovell (1998) introduced a system to segment the cellular regions within the Papanicolaou stained tissue samples, by using the active contour method. The system does not provide accurate results in counting the number of cells within the cellular regions, specially for overlapping cells. The main reason is that the active contour algorithm cannot discriminate between the contours of overlapping cells and single cells.

Hu et al. (2004) introduced an improved active model contour by using ultimate erosion and dual thresholds to isolate each cell nucleus from esophageal cell images. The system significantly improved the performance of the active contour algorithm, but still the complexity of the decision function in recognizing cell nucleus from other components of the esophageal cell, remained as the main drawback of the system.

Sadeghian et al. (2009) developed a system for segmenting nuclei and cytoplasm of cells. This system uses the Canny edge detection technique (Canny, 1986) to extract the boundaries of white blood cells. Thereafter, the system identifies nuclei using gradient vector flow (Xu and Prince, 1997) as an external and internal force for the Active counter model (Kass et al., 1988). Ultimately, the system employs Zack thresholding (Zack et al., 1977) to segment the cytoplasmic regions. The system is sensitive to the fluctuating contour of the nuclei regions which results in over-segmentation.

Nam et al. (2012) proposed a segmentation algorithm by the incorporating level set method and dual active contour modeling, to enable the segmentation of specific cells called "beta islet" in transmission electron microscopy images. The level set method includes a novel shape regularizer, to prevent over-segmentation.

The dual active contour method uses two active contours, which interact with each other, in order to avoid the short comings of the systems proposed by Hu et al. (2004) and Sadeghian et al. (2009). The proposed active contour consists of two contours initialized simultaneously inside and outside the object boundary. However, the results indicate that, although the system is insensitive to fluctuating contours, it also neglects the regions which lie deeply inside the body of cells. These regions are indicative of overlapping cells, while the system considers them as single cells. This results in low accuracy in counting the number of cells.

Li et al. (2012) introduces an automated system for segmenting cellular regions, nuclei and cytoplasm structures. This system uses the non-local means filter (Buades et al., 2011) to remove noise. Then a spatial K-means clustering algorithm is proposed to extract the initial contours of the nucleus and cytoplasm. These initial rough contours are used for the initialization of the proposed Radiating Gradient Vector Flow (RGVF) snake. Radiating Gradient Vector Flow snakes are an extension of active contours. The difference between them is that RGVF converges to boundary concavities, and unlike traditional active contour models, it does not need to be initialized close to the boundary. This significantly reduces the sensitivity of the system to fluctuating regions in the contour of cells. Moreover, the system identifies those regions which lie deep inside the body of the cells, and consequently identifies the overlapping cells. Its main drawback is high standard deviation in density inhomogeneous images.

Zeng et al. (2013) proposed an unsupervised segmentation method for extracting cell nuclei. This method segments the cell nuclei using adaptive active contour modeling based on the morphological method. First, binary cell nuclei are acquired by using an image subtraction technique. Secondly, the masks of the cell nuclei are obtained to drive adaptive region-based active contour modeling

to segment the cell nuclei. The system shows good results for density inhomogeneous images. However, it cannot detect the constituent cells of a cell clump due to the overlapping contours of the cells.

Performance of the proposed algorithms by Sadeghian et al. (2009) in segmenting cellular regions is compared with performance of the proposed algorithms in this thesis in Section 4.2.3. The results indicate that the segmentation algorithms proposed in this thesis outperform active contour techniques in segmenting cellular regions and ganglion cells within H&E stained histological images.

Post-Segmentation

As mentioned in previous sections, one of the challenges in segmenting histological structures such as different types of cells and nuclei, is the overlapping issue. Overlapping cells and nuclei reduce the accuracy of quantitative analysis. To improve the accuracy of the quantitative analysis, after segmenting the histological structures, post segmentation operation must be applied to the segmented regions. Post-segmentation recognizes the overlapping objects and decomposes them into their constituent objects. A variety of algorithms and systems have been proposed.

Watershed technique (Beucher and Lantuejoul, 1979) is widely used in different systems, but intensive over-segmentation is the main drawback. Intensity analysis approach is another type of algorithm for splitting the overlapping cells. Kumar et al. (2006) proposed a system which analyzes the cell contour and the concavity of the contour of the overlapping cells. The contours of the cells usually contain many false concave regions which creates errors in these systems.

Wang et al. (2012) and Qi et al. (2012) follow the intensity analysis approach by identifying splitting points based on the higher intensity value in the area of

overlap. These methods are sensitive to the quality of tissue staining.

Kong et al. (2011) proposed a system to split the overlapping cells by smoothing the boundary of the cells using Fourier shape descriptors (Zahn et al., 1972). It splits overlapped cells from their dominant concave region by choosing a point γ_b on the contour of the cell. It then finds γ_{b+F} and γ_{b-F} points on the cell contour which are F-points (F is set to 12) ahead and behind of γ_b. According to Kong et al. (2011), if more than 60% of the line $\overline{\gamma_{b-F}\gamma_{b+F}}$ linking γ_{b+F} and γ_{b-F} is outside the overlapped cell, then γ_b is considered to be a concave point. The system splits the clump by cutting along the two detected concave points. High sensitivity of the system to fluctuating cell contours results in high standard deviation. In Chapter 5 the results obtained by this system are compared to the results of the proposed system in this thesis. The results show that the algorithms proposed in this thesis are significantly less sensitive to the fluctuating cell contours in comparison with the algorithms in Kong et al. (2011).

2.2.2 Algorithms and Systems for Qualitative analysis

Automated qualitative analysis of cancerous tumor is the determination of a prognosis, diagnosis or treatment for cancerous tumor based on the quantitative results of histological structures and different combinations of histological regions and histological structures. Automated quantitative analysis of cancerous tissue for determining a diagnosis contains two main stages, 1) segmenting histological regions and histological structures, and 2) determining stages of the cancer using a prognosis scheme. Because different cancers have different prognosis schemes, pathologists take different approaches and analyses into consideration for diagnoses and prognoses. As a result, there is no generic algorithm for automated systems to make diagnoses or prognoses for the cancers, and different cancers

require different algorithms. Automated qualitative analysis is a highly complex approach which requires extensive analysis and hybrid algorithms.

In this section a number of algorithms and systems for grading the aggressiveness of different cancers such as neuroblastoma, prostate, mitosis staging, breast, brain and follicular lymphoma are reviewed.

Neuroblastoma Tumor

Kong et al. (2009) proposed a framework by integrating computer vision and machine learning techniques for grading the aggressiveness of NT images into undifferentiated, poorly differentiated and differentiated. Figure 2.4 indicates an overview of the system. The system has been designed for analyzing H&E stained histological images. It analyzes the tissue using a multi-scale approach which consists of multiple image resolution levels. This means that the system dynamically changes the image resolution level at which it proceeds with sequential image analysis steps. The image at each resolution level is segmented into four histological regions such as nuclei, red blood cells, neuropil and background.

To segment the histological structures, the system firstly uses a simple yet efficient threshold based method to identify the red blood cells from the tumor. The system then implements a novel segmentation method, namely Expectation Maximization Linear Discriminant Analysis (EMLDA), that uses the joint intensity-entropy domain of pixels in the L*a*b* color-space images. The proposed segmentation technique integrates the Fisher-Rao criterion (Lehmann and Casella, 1998) into the EM algorithm and clusters pixels based on their joint intensity-entropy domains in such a way that the constructed clusters can be highly distinguishable from each other. This process is repeated until the Fisher-Rao criterion, the ratio of the sum of squared between-class distances to the sum

of squared within-class distances, converges to its maxima. Thus, constituent pixels of each of the histological structures are grouped in separate clusters.

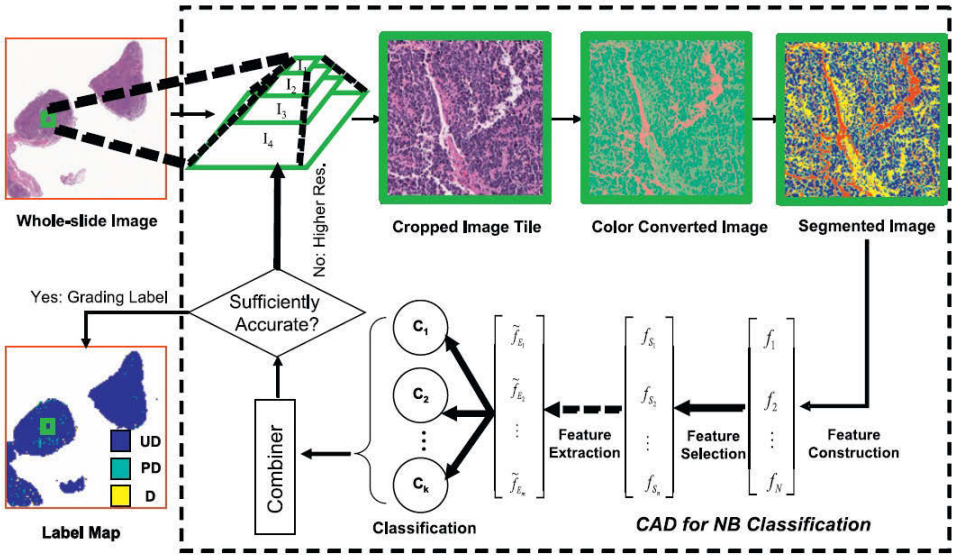


Figure 2.4: General view of the proposed system by Kong et al. (2009). The figure is taken from (Kong et al., 2009).

The main drawback of this segmentation method is that the number of classes that are generated by the clustering technique, must be set by a user according to the number of different types of histological region in each image. Therefore, the performance of the system in segmenting histological regions is highly dependent on the interaction and skill of the user in setting parameters. Moreover, in tissues with variety of different histological regions, the system classifies pixels into wrong clusters which results in segmenting the wrong types of histological regions.

The system grades NTs by combining a family of classifiers using mean of the range of intensity values, variance of the range of intensity values and homogeneity of the constituent pixels of the segmented histological regions. Base classifiers used include K-nearest neighbor, linear discriminant analysis, nearest mean, support vector machine and Bayesian. Then using a two-step mechanism,

the system combines the resulting decisions. A confidence measure is applied to each of the classification decisions to evaluate the degree of agreement across different classifiers.

The main drawback of the system is that it does not follow any particular prognosis scheme. Although it is claimed that the system used the Shimada classification, many of the essential Shimada scheme markers, such as the ratio of ganglion cells to neuroblast cells and the number of neuropil regions within the tumor tissue, are not detected and interpreted for classification of NT. The system relies on nuclei, cytoplasm and background regions only to grade the aggressiveness of NTs. Later sections of the thesis will show these omissions negatively impact on the performance of the system in classifying NTs in real life experiments. The results of the above system indicate that the mentioned issues reduce the performance of the system in providing an accurate prognosis for pathologists. In Chapter 6, the results of this system in terms of making prognoses for NTs are compared with the method proposed in this thesis.

The multi-scaling approach which is used in Kong et al. (2009) is not required for the proposed algorithms in this thesis because $20\times$ magnification is a standard magnification for analyzing the microscopic slides (Rosai et al., 2004) which shows all the necessary regions of interest. Moreover, in Kong et al. (2009) multi-scaling approach is used to eliminate unnecessary regions and focus on the regions of interest only, while in tissue spots, the image is constructed by regions of interest only.

Gland Segmentation and Prostate Staging

Jafari-Khouzani and Soltanian-Zadeh (2003) developed a system for grading prostate cancer. The system deploys multiwavelet coefficients to determine energy

and entropy of the image, then implements the K-nearest neighbor classifier to determine the grade of the prostate tissue.

Wu et al. (2005) segment human intestinal gland images using a region growing algorithm. The big, empty areas inside the intestinal glands are considered as the seeds. The algorithm then grows from the seed until the closed chain structure formed by the epithelial nuclei is covered. It then classifies the intestinal gland using neural networks.

Farjam et al. (2007) extract the gland from prostate tissue using texture features for the gland components such as stroma, lumen and nuclei. The system then exploits the K-means clustering technique to group the components into three clusters. Finally, the system grades the gland according to an analysis of the ratios of the components.

Naik et al. (2007) proposed a system for grading prostate cancer by segmenting nuclei and glands within the tissue. The system uses a trained Bayesian classifier to detect the glands and a level-set algorithm is employed to recognize the glands among the other histological structures within the tissue. The system then grades the malignancy of the prostate cancer by analyzing the morphological features of the segmented regions. The system implements a Support Vector Machine (SVM) to predict the corresponding malignancy grade tissue based on the identified morphological features.

For a number of reasons, the above systems are not applicable to NTs. First is the slides of NTs are usually stained by the H&E method which provides noisy images for computerized analysis, and the system must be insensitive to that noise, otherwise it significantly reduces the accuracy of the system in segmentation process. Additionally, most NTs contain extremely dense cellular regions which demand robust algorithms that can accurately locate and segment them

within the tumor. Finally, the morphological features of the histological structures within the NT are different to those of prostate cancer, and there is no gland structure within the tissue of NTs.

Mitosis Staging

The process of dividing the nucleus of a cell into two nuclei for generating a new cell is called mitosis. The ratio of mitotic cells to normal cells and the phases of mitosis such as Interphase, Prophase, Prometaphase, Metaphase, Anaphase1, Anaphase2, and Telophase are markers for determining the aggressiveness of many cancerous tumors.

Harder et al. (2006) introduced an approach to automatically segment, track, and classify a cell's nuclei mitotic phases. For segmenting the cellular regions, the system implements adaptive thresholding technique. They then develop a tracking scheme to observe the splitting of cells during mitosis. Based on the obtained tracking results, the system automatically selects the most informative region of the image for feature extraction. Finally, the system uses SVM to classify the cells into the seven mitotic phases.

Zhou and Li (2009) proposed a cell segmentation method to identify the different nucleic mitosis phases. The system applies the watershed technique to the cells in different mitotic stages, and uses a novel hybrid merging algorithm to reduce the negative effects of high over-segmentation derived by the watershed technique. This algorithm combines the compactness score with the probability distribution function score, to merge the fragmented cells produced by the watershed technique. Finally, by utilizing context information of time-lapse data and a Markov model, the mitosis phases of the nuclei are determined.

Although the above-stated techniques are very accurate in segmenting the

nuclei regions, there are not applicable for NTs, because NT contains a specific type of cell which is called a ganglion cell. This cell is composed of nucleoli, nuclei and cytoplasm regions which require a specific algorithm for segmentation, while the above-stated algorithms consider nuclei regions of the cells only.

Breast Cancer

Weyn et al. (1998) developed a system which implements wavelets for grading breast cancer in multi-scale image. The system determines parameters for describing the texture of chromatin in the nuclei regions. The parameters are calculated by comparing the performance of co-occurrence, densitometric, and morphometric parameters using automated K-Nearest Neighbor (KNN) classification.

Lee and Street (2003) proposed a system for automated localization, segmentation, and the grading of breast cancer nuclei using a neural network. The proposed neural network learns from the cluster shapes and it grows by creating a new cluster based on the previously identified clusters. All the nodes in the hidden layer of the neural network system represents a cluster, and they are used as a template to provide accurate nuclei localization and segmentation.

Naik et al. (2008) extended their work in (Naik et al., 2007) to locate and segment the nuclei regions within breast tissue for staging the aggressiveness of the cancer. The system applies low level information based on pixel values, high level information based on relationships between pixels, and domain specific information to determine the stage of the cancer.

Boucheron et al. (2010) proposed a system for staging the grade of breast cancers within the images of H&E stained tissues. The system integrates the marker-based watershed segmentation technique with the concavity-based seg-

mentation technique to extract the cellular and nuclei regions. Ultimately the system classifies the stage of breast cancer using a set of object-level features and a linear classification.

The grading of breast cancer is done by analyzing particular types of cells. The morphological properties of those cells are different to those of the cells in NTs. This means that different segmentation approaches must be taken into consideration for locating and extracting each of the cells. Moreover, in NTs the ratios of the different histological regions such as neuropil regions and cellular regions are used as a marker for determining the stage of the tumor. Finally, the grading scheme for breast cancer is different to the grading scheme of NTs.

Brain Cancer

Reddick et al. (1997) introduced a hybrid neural network method by integrating the Kohonen self-organizing neural network for segmentation of the histological structures of interest with a multilayer back-propagation neural network for classifying the grade of the tumor. To distinguish different tissue types from each other, the system uses volumetric measurements of brain structures, relative to intracranial volume.

Ho et al. (2002) developed a new method for the automated segmentation of histological structures from brain tumor medical images. The method implements an intensity-based fuzzy rule to classify the images into tumor and background sections. The system uses the obtained tumor probability map to locally guide the propagation direction and speed of a level-set. Moreover, the obtained probability map of the tumor is exploited to derive an automatic initialization of the level-set. Finally the system performs a comparison between the pre and post contrast difference images to classify the image into background and tumor regions.

Demir and Yener (2005) proposed a system that automatically grades the aggressiveness of the brain cancer tumors. The system follows three major steps:

1. Noise reduction to improve the focal area determination success. The output of this step is the localizing and segmenting of nucleus/cell regions.

2. Representing the focal areas that provide distinctive objective measures.

3. Measuring the aggressiveness of the tumor using machine learning approaches such as fuzzy systems and logistic regression.

Although the tissues of brain tumors are similar to the tissues of NTs because both contain the same types of cells (called neural cells), there are huge differences between their prognosis schemes, histological regions and histological structures. Different morphological and classification criteria between NTs and brain tumors demand different and separate algorithms for each of the tumors.

Follicular Lymphoma Grading

Sertel et al. (2009) developed a CAD system that for assisting pathologists in the grading of Follicular Lymphoma. The system implements a novel approach for segmenting histological structures within the images, by incorporating model based intermediate representation and low level color texture analysis. To this end, they first identify basic histological structures in the image and model the connected components of such regions using ellipses. A set with a wide range of features is constructed from this intermediate representation to characterize the tissue. Using this representation, the relative amounts and spatial distributions of these histological components are measured to determine the grade of the follicular lymphoma tumor. Since the tissue samples of low grades are better

characterized by low level color texture features, the system deploys non-linear color quantification based on a color texture to classify the low grades of follicular lymphoma.

Cooper et al. (2009) developed a registration approach based on the matching of small, salient, anatomical features. Most tissues contain small features such as blood vessels which have approximately common morphological properties. This makes the process of localization and segmentation easier. They propose a two-stage algorithm to localize and segment these features: first rigidly, to roughly align the images, then non-rigidly, to correct for elastic distortions introduced by preparation. The first stage uses mismatch-tolerant voting procedure (Huang et al., 2006). With the rough alignment of the images calculated, the second stage deploys coherent local networks of matched features between the images to improve the matching degree, and to reduce the probability of mismatch.

Both of the above mentioned methods for grading Follicular Lymphoma are based on the morphological properties of the diagnostic features. The morphological features of the Follicular Lymphoma are different in terms of shape, size and color to the morphological features of NTs. The proposed texture analysis algorithm in the above systems can provide a high performance in locating the histological structures with an outstanding texture pattern. However, relying on the texture of NTs histological structures reduces the performance of the system because of those fluctuating quality of the H&E staining method and the nuclear texture of histological structures.

2.3 Research Gaps

In summary, to develop a CAD system for the quantitative and qualitative analysis of NTs and for determining a diagnosis, prognosis and treatment, two main steps must be taken into consideration: 1) segmenting histological regions and histological structures, 2) grading the cancer based on its prognosis scheme. As reported in the literature review most the common issues among the developed CAD systems are 1) the wide range of intensity variation in the histological images, 2) developing an appropriate segmentation algorithm for locating and extracting the histological regions and histological structures of interests, 3) dealing with overlapping cells or nuclei to improve the performance of quantitative analysis, and 4) developing accurate computerized rules based on the prognosis scheme of the tumor to determine the correct prognosis for cancers.

In case of NTs a small number of systems are proposed for segmenting histological regions and histological structures of the tumor such as Gurcan et al. (2006), Gurcan et al. (2007) and Ruiz et al. (2008). The main issue is that these systems are tailored to extract a limited number of histological regions or histological structures such as cellular regions or nuclei structures, while as mentioned in Section 2.1.1, NTs contain variety of histological regions and histological structures to be analyzed. Moreover, the issue with most of the systems is that they do not integrate quantitative and qualitative analysis. As explained in section 2.2.2, Kong et al. (2009) proposed the only CAD system for determining a prognosis for NTs. However, the system does not consider all of the necessary histological regions and histological structures of NTs, and it does not follow a prognosis scheme to determine a prognosis for NT.

In this thesis, a CAD system is proposed to determine a prognosis for NTs

based on the Shimada prognosis scheme for extracting all of the necessary histological regions and histological structures of NTs. Moreover, the system addresses all the above mentioned gaps in the segmentation and decision making process. In Section 6.4 the results of experiments indicate that the system proposed in this thesis outperforms the system proposed by Kong et al. (2009) in terms of grading NTs.

Finally each cancer has its own histological regions and histological structures with their specific morphological features and characteristics. Thus, to detect each of those specific histological regions or histological structures, an algorithm which is tailored to locate and segment that region must be deployed. For testing the performance of the proposed system in this thesis, the most relevant systems are taken into consideration. As most of those systems are not fully tailored to extract the specific histological regions and histological structures of NTs such as ganglion cells, neuroblast cells and neuropil regions, and as they require some changes in their algorithms such as tuning parameters, they provide a lower performance in comparison with the algorithms proposed in this thesis.

Chapter 3

Model of the CAD System

3.1 Stages of the Model

As explained in Section 2.1.1, NTs contain a number of histological regions such as cellular regions and neuropil regions, and histological structures such as ganglion cells, neuroblast cells, nuclei and cytoplasm. Clinicians use prognosis schemes to analyze the ratio and relationship between the above histological regions and structures to determine the diagnosis, prognosis and treatment for NTs. Thus, the proposed CAD system must firstly identify and extract those regions from the histological regions and then apply computerized rules based on the histopathology scheme for making prognosis. To this end, four main stages must be taken into consideration: 1) Pre-segmentation to enhance the quality of the histological images for obtaining greater accuracy in segmenting histological regions and histological structures, 2) segmentation to locate and extract histological regions and histological structures within the tumor image, 3) post-segmentation to deal with the issue of overlapping cells and nuclei, and 4) grading NTs based on the Shimada scheme. Figure 3.1 indicates each of the stages.

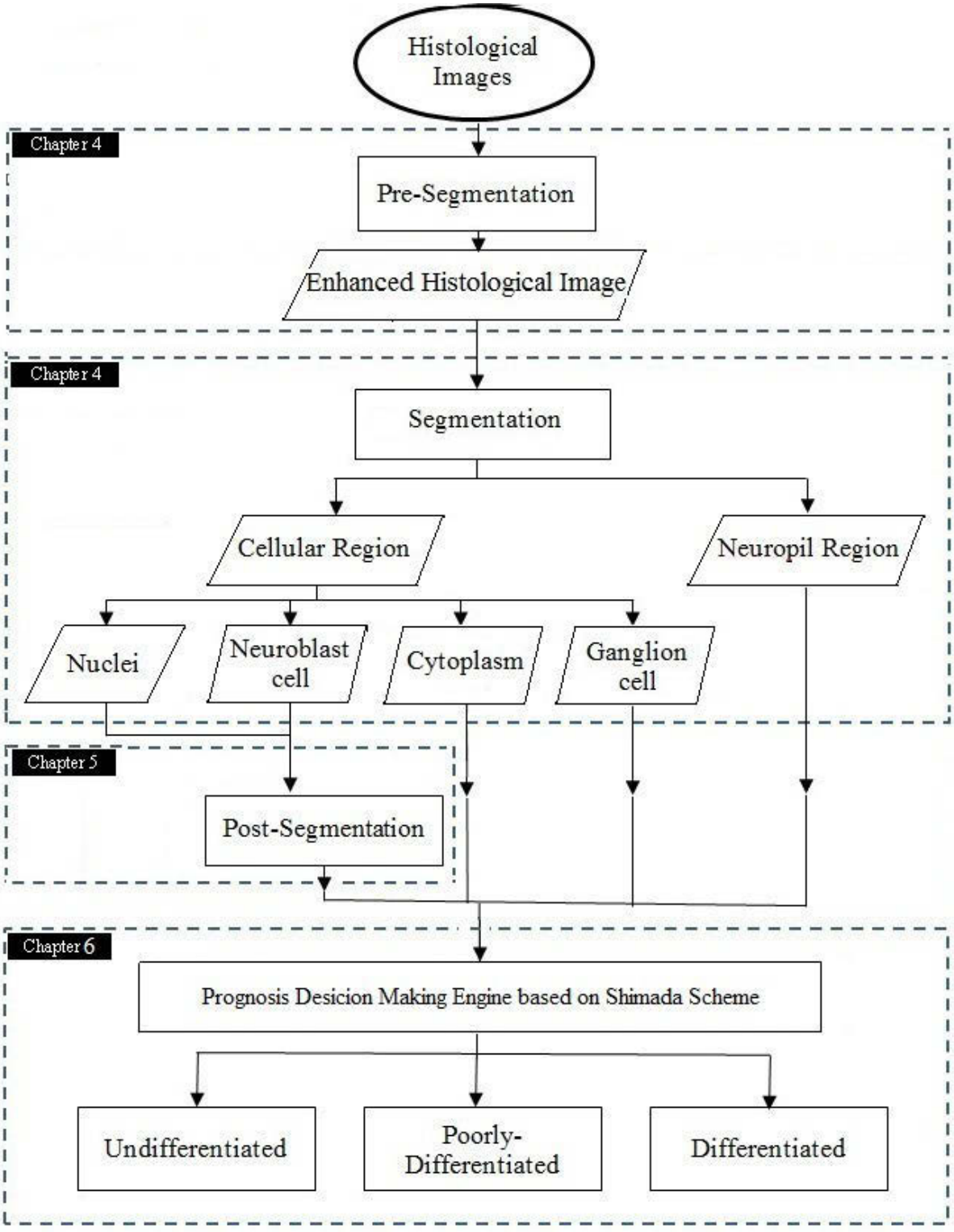


Figure 3.1: Thesis Overview.

In the pre-segmentation process, to improve the low quality of the H&E stained histological images for further computerized analysis, novel global and local intensity reduction variation techniques are proposed. Section 4.1 proposes and validates pre-segmentation algorithms.

In the segmentation process, in order to identify the numerical patterns of different types of histological regions and histological structures, a series of algorithms are proposed. These algorithms segment cellular regions, neuropil regions, ganglion cells and neuroblast cells by applying the developed histological filtering, centroid mapping, and depth analysis algorithms. In Section 4.2 all the proposed algorithms for segmenting the above histological regions and histological structures are described and validated.

In the post-segmentation process, to overcome problems with overlapping objects, two algorithms are proposed for improving the performance of the system in automated quantitative cellular analysis. The first aim of post-segmentation is to distinguish overlapping cells from single cells by using developed algorithms such as concavity dominance, diameter equality, and cell area analysis. The second aim is to split overlapping cells into single cells by utilizing the introduced concave hull critical point determination, convex region decomposition and the shortest path determination algorithms. In Chapter 5 the proposed algorithms for post-segmentation are introduced and validated.

For grading the aggressiveness of NTs, a prognosis decision engine is proposed by establishing a set of computerized rules according to the Shimada classification scheme for analyzing the segmented histological regions and structures. This section is developed in close consultation with pathologists from the department of histopathology at CHW. In Chapter 6 the developed prognosis decision engine is described and validated.

3.2 The Overall Research Strategy

To achieve the research objective the following procedures are taken into consideration during the development and design of the CAD system.

3.2.1 Awareness of Problems

The first step is to analyze the limitations of the existing algorithms and systems, and acknowledge the significant research problems. The research problems indicate a gap between existing applications and the expected outcomes. Research problems in the domain of digital pathology can be determined by investigating different sources such as clinical experiences, literature reviews and observation of practical applications, existing CAD systems, their limitations and range of dissemination. A clear definition of the research problem provides a precise roadmap for the research throughout the process of development (Turban et al., 2007).

3.2.2 Suggestions

After determining the research problems, a tentative design must be produced. The tentative design describes the nature and method of development of prospective artifacts. Suggestion is a creative process during which new concepts, models and functions of artifacts are demonstrated.

3.2.3 Development

New algorithms and systems are developed based on the suggested tentative design. Novelty is a principle in the process of designing and developing a new algorithm or systems. Thus, the developed algorithms and systems must propose a new technique, models or approach to address the problems within the domain.

The development process is often iterative, where an initial prototype is first built and then evolves as the researcher gains deeper comprehension of the research problems.

3.2.4 Evaluation

The performance of the developed algorithms and systems must be evaluated according to defined criteria in the research proposal and the suggested design. Based on the obtained results by evaluation, developed algorithms or systems may or may not meet expectations. In case of unsatisfactory results, a revision of the suggestion and development stages must be considered to improve the results.

3.2.5 Conclusion

Conclusion is the final stage of the research process. Most often, it is the result of satisfying of the evaluation of the algorithms and systems which have been developed. While there may still be some discord between expectations in the proposal and the outcome of the research, but the research design concludes when the results of evaluation are considered to be satisfactory.

3.3 Data Acquisition, Deployed Software and System Validation

All images of neuroblastoma tumors used in this research were provided by the Tumor Bank of The Kid's Research Institute at The CHW which is a facility approved by the CHW Human Research Ethics Committee to oversee the provision of human tumor tissue and related data for research investigations. The Tumor

Bank is compliant with the local policy, national legislation and ethical mandates for use of human tissue in research which are in keeping with the Declaration of Helsinki. Patient consent has been obtained for all samples stored in the Tumour Bank which have been collected during the routine course of clinical management and histopathological assessment for their disease, and are de-identified for this study such that no patient specific details were provided.

The dataset contains multiple tissue arrays which are scanned by the Aperio ScanScope system. Figure 3.2 indicates a tissue array. The tissue arrays are analyzed under 20× magnifications and Aperio ScanScope scanned them at a resolution of 0.46 microns per pixel. The size of each tissue array is 2.5 GB on average. The tissue spots on the tissue array are taken from different NTs with different stages. In total, 15 tissue arrays each with approximately 50 tissue spots are used for training and testing the developed algorithms. Each of the tissue spots was cropped to an image of size 512×512 pixels with JPEG format and compression rate of 13:1. The test set and the training set for each of the developed algorithms in each of the chapter were different, and different tissue arrays and tissue spot images were used for testing and training the segmentation algorithms, post-segmentation algorithms and classification algorithms.

All of the results obtained by the developed systems in this thesis were validated by a pathologist from the department of histopathology in the Children's Hospital at Westmead, Sydney. The pathologist was considered as the ground truth and this is the baseline for all of the validation in all chapters. To measure the performance of the proposed algorithms in this thesis precision, recall and F-measure (Powers, 2011) of the obtained results by the algorithms are computed by comparing them with those of the pathologist. This means that manual quantitative and qualitative analysis were performed by the pathologist and the

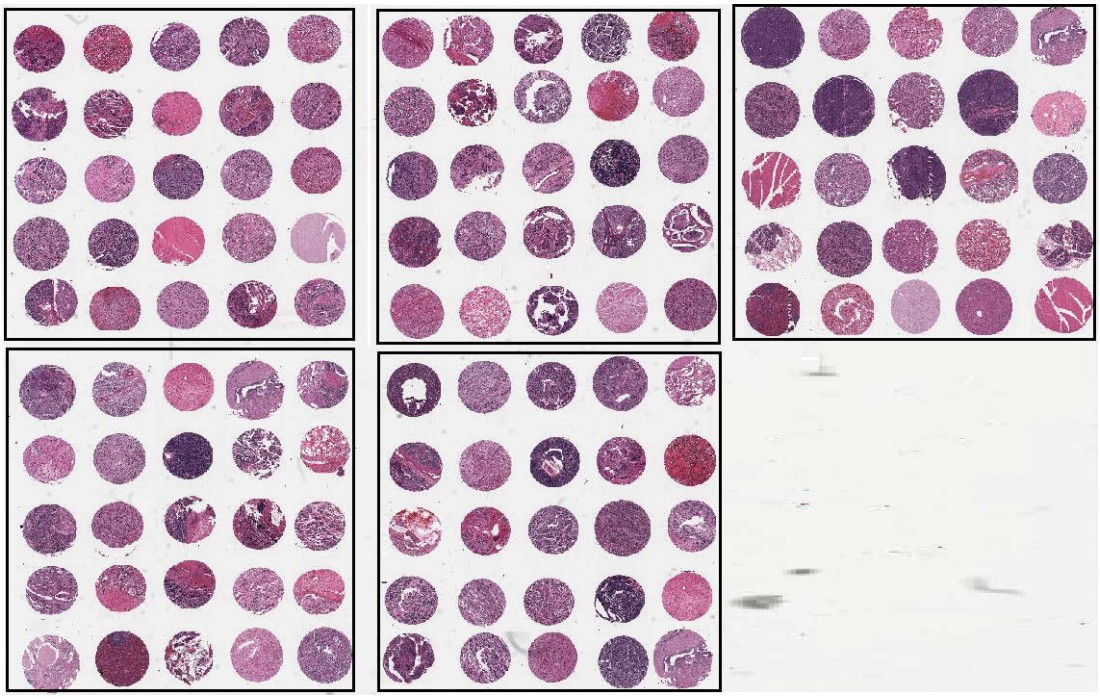


Figure 3.2: An example of a tissue array that is used on this thesis.(1× magnification).

results used as the ground truth for measuring the performance of the CAD. As mentioned in Section 2.1, NT is a paediatric cancer, and the reviewer pathologist specializes in paediatric cancers, an important factor in the accuracy of the validation. According to Tomaszewski et al. (2000) the concordance between the diagnosis and prognosis of specialized pathologists and central pathologists is significantly higher than the diagnosis and prognosis concordance between general pathologists and central pathologists. Finally, a recent audit of the department of histopathology at CHW shows 0% historical discordance in classifying NTs.

The developed algorithms for the CAD system use a number of parameters which are required to be set through relevant training program. This means that the performance of the algorithms are evaluated using different values of the parameters, and those values are chosen for the parameters which provide the

highest performance for the algorithms. The performance of each algorithm is evaluated by the pathologist.

All the algorithms were developed using MATLAB (The MathWorks, Inc., Natick, MA) and experiments were run on a computer with 2×3.47 GHz processors and 12 GB RAM.

Chapter 4

Segmenting Histological Regions and Histological Structures in H&E Stained Histological Images

This chapter addresses **Contribution 1** and **Contribution 2** listed in Section 1.5 by: 1) proposing global and local intensity variation reduction approaches to enhance the quality of H&E histological images, and 2) introducing a number of hybrid segmentation algorithms for identifying and extracting different histological regions and histological structures within NT.

Contribution 1 is motivated from the fact that pathologists identify different types of histological regions under the microscope based on their color. The images that are used in this thesis are derived from H&E stained tissue glass slides. The main drawback of H&E for CADs is the fluctuating quality of staining. For example, similar histological structures, such as cellular regions in different locations of the slide, are stained with different f of blue. The colour variation reduces the performance of the CAD in detecting cellular structures.

Contribution 2 develops algorithms for segmenting histological regions. Those considered for segmentation are cellular regions and neuropil regions, and histological structures such as nuclei, cytoplasm, ganglion cells and neuroblast cells. All the above stated histological regions are used by the Shimada classification scheme to grade NTs Shimada et al. (1999). Figure 4.1 shows the differences and morphology of those histological structures in NT.

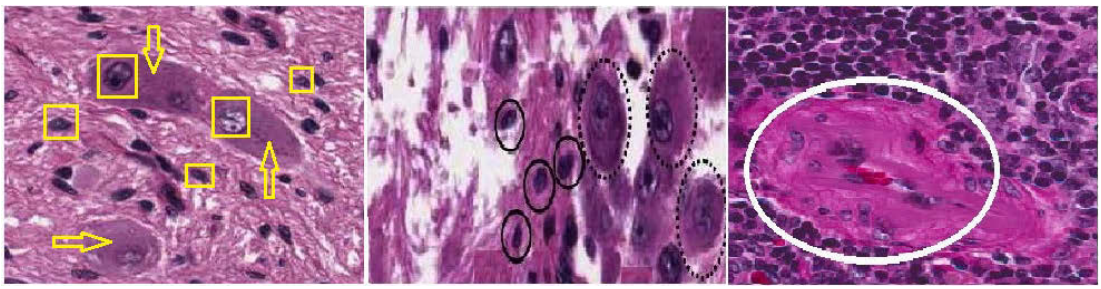


Figure 4.1: Cytoplasms are indicated by arrows, nuclei are bounded by rectangles, ganglion cells are shown by dashed ellipses, neuroblast cells are shown by back circles, and neuropile region is indicated by white ellipse.

This chapter contains three sections, 1) pre-segmentation which explains the proposed algorithm for reducing the intensity variation within the histological images of NTs, 2) segmenting cellular regions, neuropil regions, nuclei, cytoplasm and ganglion cells, and 3) segmenting neuroblast cells.

4.1 Pre-Segmentation

Pathologists identify different types of histological regions under the microscope based on their color. Similar histological regions in different locations of the tissue have the same color but they are stained with different shades of the color. The human brain is able to deal with this color variation. However, in computerized analysis different shades of a color mean different intensity ranges, and conse-

quently different histological regions. Therefore, a wide range of intensity values for similar histological structures along the image, reduces the performance of the system in the segmentation process. As a result, the first step in automated analysis of cancerous tumors is to reduce the color variation between similar histological structures.

This section addresses Contribution 1 of this thesis. The proposed algorithm enhances the fluctuating quality of the histological images by reducing the intensity variation within the images. Two algorithms are proposed to address the issue, global intensity variation reduction and local intensity variation reduction.

4.1.1 Methodology

Global and local intensity variation reduction techniques are proposed for the pre-segmentation process. The proposed global intensity variation reduction technique is based on a color quantization technique (Gonzalez et al., 2004) which is applied to the entire image. The local intensity variation reduction technique transforms the color-space of the images into a perceptually uniform color-space, partitions the images into several mosaics by grouping homogeneous pixels in terms of joint intensity-spatial domain, and unifies the intensity range of each mosaic. This reduces the intensity variations between the constituent pixels of the same types of histological regions. Figure 4.2 indicates the schematic view of the proposed pre-segmentation approach.

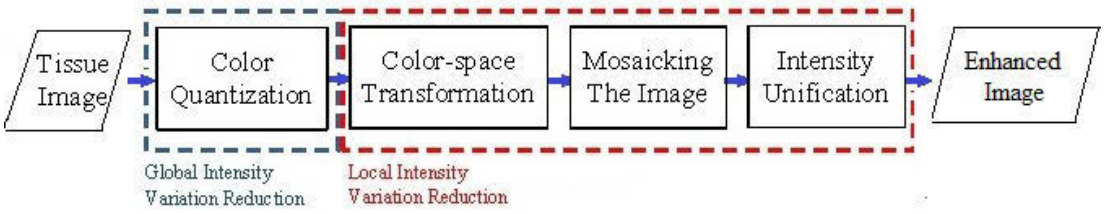


Figure 4.2: Schematic view of the proposed pre-segmentation algorithm

4.1.2 Proposed Approach

Global Intensity Variation Reduction

To reduce the wide range of intensities in the entire image, the smallest RGB color cube which contains all the existing colors of the original image is found. Then the colors along the longest axis of the color cube are sorted using a median cut algorithm (Heckbert, 1982), which divides the longest axis of the color cube into two boxes at the median point. Median cut sorts data into a series of sets and cuts each set at the median point. This process is repeated until the color cube has been divided into ξ boxes. Each of the boxes contains an approximately similar range of colors. All the colors falling within each box are mapped to the color value at the center of the box. In this thesis $\xi = 12$ is set empirically using the approaches described in Section 3.3 on 25 training images. This means that each image contains 12 colors only. Figure 4.3b clearly indicates an image after the application of the color quantization technique, showing that it contains a lesser number of colors in comparison with the original image.

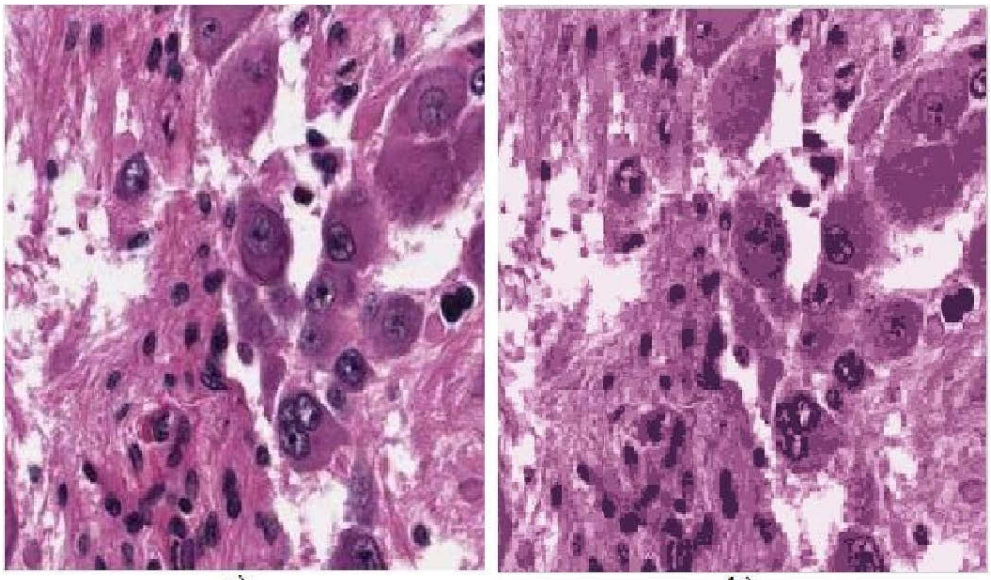


Figure 4.3: a) Original image, (b) global intensity variation reduction.

Local Intensity Variation Reduction

The next step is to reduce the intensity variation between the constituent pixels of homogeneous regions. To achieve this, the homogeneous pixels of the image are grouped, and the intensity variation between them is reduced. The proposed algorithm for local intensity variation reduction in this thesis contains three stages as follows:

1) Color-space Transformation: The initial color-space of the images is RGB, which is not perceptually uniform. It means that there is no mathematical pattern between different colors. In analyzing the images of the tumor tissues, pixels are analyzed and they are compared to each other. To obtain accurate results they must be analyzed in a robust and consistent color-space. As a result the RGB is transformed to L*u*v* as a uniform color-space which describes the differences between the colors using Euclidean distance. L*u*v* color space is

calculated by (Pratt, 1991),

$$L* = \begin{cases} (\frac{29}{3})^3 \times \frac{V}{V_i}, & \frac{V}{V_i} \leq (\frac{6}{29})^3 \\ 116 \times (\frac{V}{V_i})^{\frac{1}{3}} - 16 & \frac{V}{V_i} > (\frac{6}{29})^3 \end{cases} \tag{4.1}$$

$$u* = 13 \times L* \times (u' - {u'}_n)$$

$$v* = 13 \times L* \times (v' - {v'}_n)$$

where V_i is luminance of pixel i, and V is luminance of white point respectively, ${u'}_i$ and ${v'}_i$ are the chromaticity coordinates of pixel i, and u' and v' are chromaticity channels of white pixel.

2) Mosaicking the Image: To group the pixels, each image is considered as a two dimensional lattice of three dimensional vectors. The lattice indicates the spatial domain of the pixel in the image and the vectors show the intensity of the pixels. Thus, the histological image is considered as a feature-space with five features where the intensity value of the pixels provides three features, namely, L^*, u^* and v^*, and the spatial domain supplies two features being the x and y coordinates of the pixels within the image. It means that the pixels are grouped based on the joint intensity-spatial domain. To find the homogeneous pixels, mean-shift (Comaniciu and Meer, 2002) is used as a density estimation technique which regards the feature-space as a probability density function, and finds the modes of the dense regions. Because color image is a multivariate feature space, a multivariate kernel density estimator is required to be implemented for the mean-shift algorithm. The mean-shift procedure explained in Section 2.2.1 can be applied with a number of modifications as described in follows:

The kernel $K(X)$ for color image in L*u*v* space is defined as

$$K_{h_S,h_I}(X) = \frac{C}{h_S^2 h_I^3} k\left(\left\|\frac{X^S}{h_S}\right\|\right)^2 k\left(\left\|\frac{X^I}{h_I}\right\|\right)^2 \tag{4.2}$$

where C is the normalization constant, h_S is spatial bandwidth, h_I is intensity bandwidth, X^S is spatial domain of pixel X, X^I is the intensity value of a pixel in 3-d vector, and k is a function which is called the profile of the kernel. A well-known density estimator using (4.2) as the kernel is

$$\hat{f}_{h_S,h_I}(X) = \frac{1}{nh_S^2 h_I^3} \Sigma_{i=1}^{n} k\left(\frac{X - X_i}{h}\right) \tag{4.3}$$

where X_i is the i^{th} pixel in the image, and n is the total number of pixels. By employing the profile function of the kernel, the density estimator (4.3) becomes

$$\hat{f}_{h_S,h_I,K}(X) = \frac{C_{k,2,3}}{nh_S^2 h_I^3} \Sigma_{i=1}^{n} k\left(\left\|\frac{X^S - X_i^S}{h_S}\right\|\right)^2 k\left(\left\|\frac{X^I - X_i^I}{h_I}\right\|\right)^2 \tag{4.4}$$

As explained earlier mean-shift identifies the modes of density which lie where the gradient is zero, thus using (4.4), the density gradient estimator is defined as

$$\frac{\hat{f}_{h_S,h_I,K}(X)}{\sigma X^S X^I} = \frac{4C_{k,2,3}}{nh_S^6 h_I^7} \Sigma_{i=1}^{n} (X^S - X_i^S)(X^I - X_i^I) k'\left\|\left(\frac{X^S - X_i^S}{h_S}\right)\right\|^2 k'\left\|\left(\frac{X^I - X_i^I}{h_I}\right)\right\|^2 \tag{4.5}$$

by defining

$$g(X) = -k'(X) \tag{4.6}$$

new kernel $G(X)$ is derived as

$$G(X) = C_{g,2,3} g\left(\left\|\frac{X^S}{h^S}\right\|^2\right) g\left(\left\|\frac{X^I}{h^I}\right\|^2\right) \tag{4.7}$$

where $c_{g,2,3}$ is the normalization constant. Substituting $g(X)$ into (4.5) yields

$$\frac{\hat{f}_{h_S,h_I,K}(X)}{\sigma X^S X^I} = \frac{4C_{k,2,3}}{nh_S^6 h_I^7}\Sigma_{i=1}^{n}(X^S - X_i^S)(X^I - X_i^I)g\left\|\left(\frac{X^S - X_i^S}{h_S}\right)\right\|^2 g\left\|\left(\frac{X^I - X_i^I}{h_I}\right)\right\|^2 =$$

$$\frac{4C_{k,2,3}}{nh_S^6 h_I^7}\left[g\left\|\left(\frac{X^S - X_i^S}{h_S}\right)\right\|^2 g\left\|\left(\frac{X^I - X_i^I}{h_I}\right)\right\|^2\right]\times$$

$$\left[\frac{\Sigma_{i=1}^{n}(X^S X_i^I - X^I X_i^S - X_i^S X_i^I)g\|(\frac{X^S - X_i^S}{h_S})\|^2 g\|(\frac{X^I - X_i^I}{h_I})\|^2}{\Sigma_{i=1}^{n} g\|(\frac{X^S - X_i^S}{h_S})\|^2 g\|(\frac{X^I - X_i^I}{h_I})\|^2}\right] - X^S X^I \tag{4.8}$$

where $X^S X^I$ is the center of the kernel density estimator window. In (4.8) the first term is the kernel density estimator

$$\frac{\hat{f}_{h_S,h_I,G}(X)}{\sigma X^S X^I} = \frac{4c_{k,2,3}}{nh_S^6 h_I^7}\left[g\left(\left\|\left(\frac{X^S - X_i^S}{h_S}\right)\right\|^2\right) g\left(\left\|\left(\frac{X^I - X_i^I}{h_I}\right)\right\|^2\right)\right] \tag{4.9}$$

and the second term is the mean-shift

$$m_{h^S,h^I,G}(X) = \left[\frac{\Sigma_{i=1}^{n}(X^S X_i^I - X^I X_i^S - X_i^S X_i^I)(g\|(\frac{X^S - X_i^S}{h_S})\|^2)(g\|(\frac{X^I - X_i^I}{h_I})\|^2)}{\Sigma_{i=1}^{n} g(\|(\frac{X^S - X_i^S}{h_S})\|^2)g(\|(\frac{X^I - X_i^I}{h_I})\|^2)}\right] \tag{4.10}$$

In the next step, a unique label is assigned to each mode, and consequently all the constituent pixels receive the same label as their mode. All the pixels with identical labels form a mosaic as shown in Figure 4.4b thus, a mosaicked image is a set of matrices as

$$\theta = \mathbf{M}_a, \quad \forall a \in \{1, ..., B\} \tag{4.11}$$

where $\mathbf{M}$ is a $d \times 3$ matrix, d is the total number of pixels within mosaic a, B is the total number of constructed mosaics and columns represent L*, u*, and v* values.

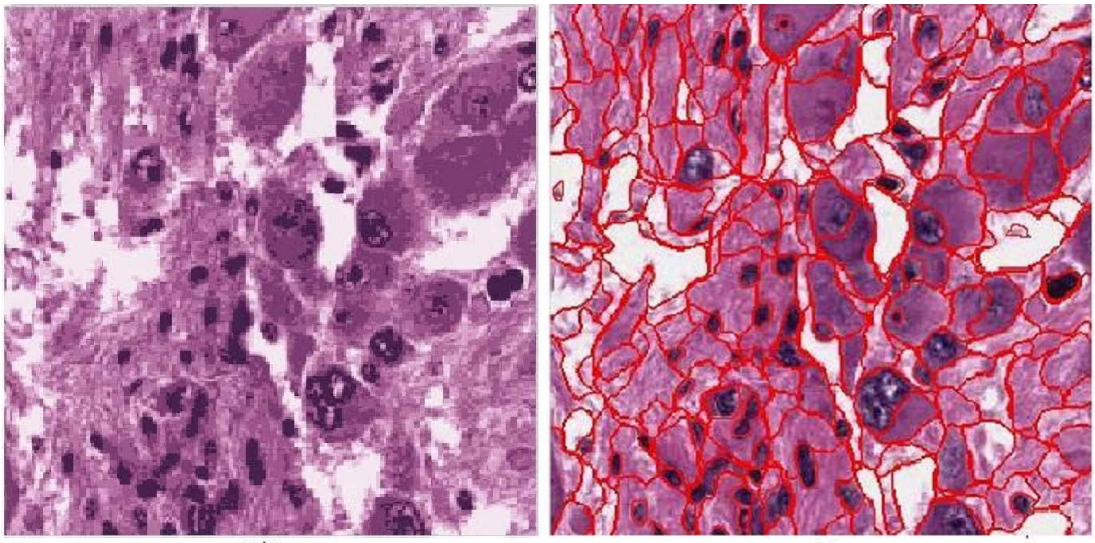


Figure 4.4: a) The image after global intensity unification, (b) the image after mosaicking.

3) Intensity Unification: Although the constituent pixels of a mosaic are homogeneous, there is a small variation between them in terms of intensity. This must be removed because each mosaic belongs to one type of histological region only, and the histological regions of the same type are stained with one color. Consequently different shades of the color should not be taken into consideration. Thus, each mosaic can be represented by one color only instead of different shades of a color.

The color of each mosaic is unified by first obtaining the mean intensity value of all pixels within each of the mosaics. The second step involves assigning the mean intensity value of each of the mosaics to their corresponding constituent pixels. This thesis denotes a unified intensity mosaic by $\bar{\mathbf{M}}_a$ which is a row vector of length 3 where the columns represent the mean values of the L^*, u^* and v^* of the matrix $\mathbf{M}_a$. Figure 4.5b shows an image after applying the local intensity variation reduction. Validation for the pre-segmentation is done as part

of validation of the segmentation.

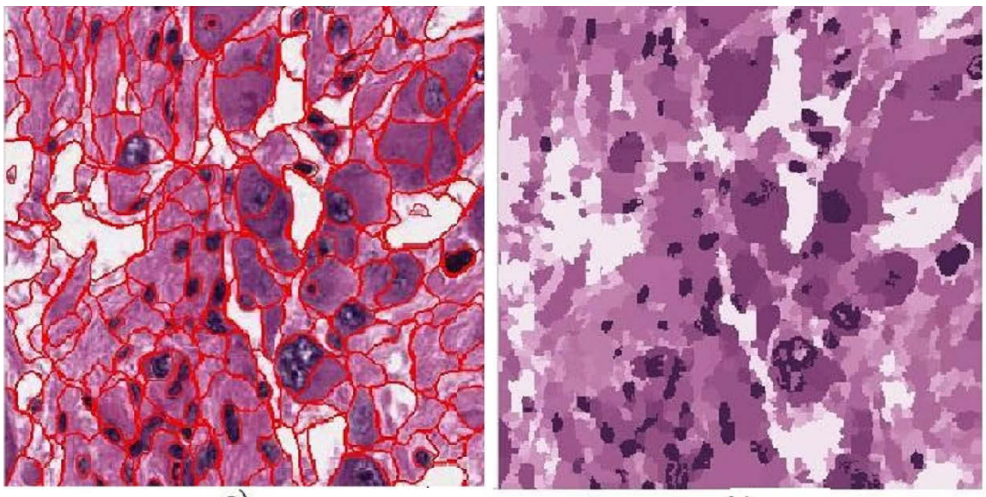


Figure 4.5: a) The mosaicked image, b) the image after intensity unification

4.2 Segmenting Cellular Regions, Neuropil Regions, Ganglion Cells and Cytoplasm Components

Making a prognosis for cancer requires an intensive quantitative analysis of cell components within tissue biopsy. One important criterion in the prognosis of NTs is the ratio of ganglion cells as mature nervous cells to neuroblastic cells as immature nervous cells (Shimada et al., 1999).

A major issue for developing the above system is extracting neuropil regions and discriminating ganglion cells from other histological structures within segmented cellular regions. The main challenge for neuropil regions is that they do not follow a specific pattern in terms of morphology. Color is the only marker which is used by pathologists to identify them, and the main issue in discriminat-

ing ganglion cells from other cellular components such as cytoplasm, nuclei and neuroblast cells is that all of them have some degree of similarity in terms of morphology and color. The specific morphological features of ganglion cells do allow computerized analysis for distinguishing them from other cellular components. The main morphological properties of a ganglion cell are: a huge cytoplasmic region around its nucleus, and the size of a ganglion cell is twice the size of its nucleus (Shimada et al., 1999). Figure 4.6 indicates an H&E stained tissue of NT containing ganglion cells, neuroblastic cells and neuropil regions.

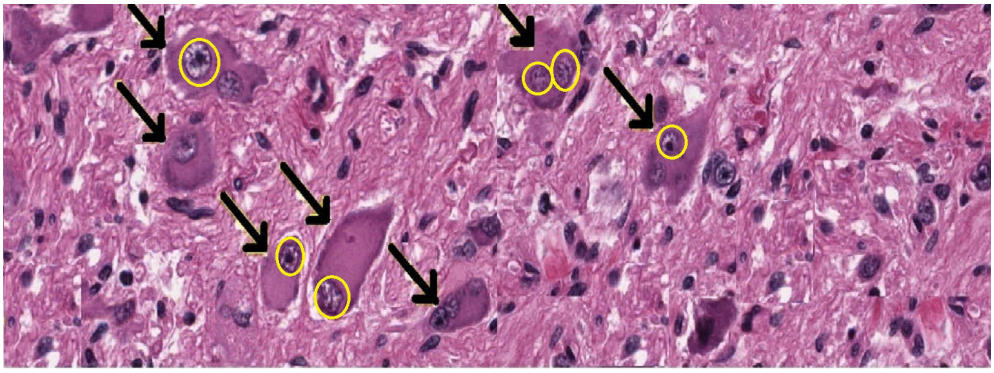


Figure 4.6: The circles shows cell nuclei, the arrows indicate ganglion cells with their nuclei and cytoplasm, the dark blue spots are neuroblast cells, and the pink regions among the cells are neuropil regions.

This section addresses the second contribution of this thesis by extracting the neuropil and cellular regions, and decomposing cells into their nucleic and cytoplasmic components for discriminating between ganglion cells and other cellular components. To this end, the proposed pre-segmentation algorithm is applied to the images for enhancing their quality and mosaicking the images. Then a novel segmentation algorithm is proposed as an histological filter. It filters the mosaics which are associated with neuropil and cellular regions by measuring the degree of similarity between the mosaics and the regions of interests in terms of intensity. Lastly, a centroid mapping and depth analysis technique is proposed

to discriminate ganglion cells from other cellular components by analyzing the cytoplasmic regions around the nuclei.

4.2.1 Methodology

For segmenting the cellular regions and neuropil regions, histological filters are developed. Then, to extract cytoplasm and ganglion cells from the segmented cellular regions a centroid mapping and depth analysis technique is proposed. Figure 4.7 indicates an overview of the proposed system. To develop and test the algorithms, 50 tissue spots from 10 different tissue arrays are taken. Different tissues are derived from different tumors. The test set contains 25 images and the training set has the other 25 images. The images of test and training sets are different from each other.

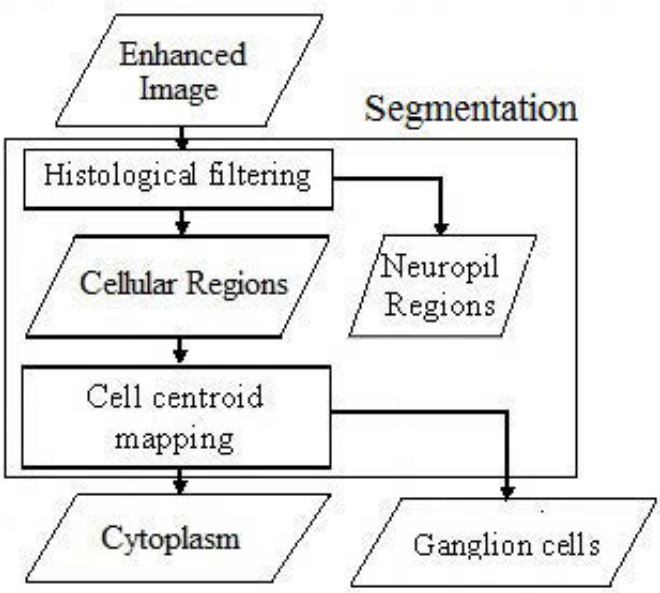


Figure 4.7: Schematic methodology of the proposed system to extract ganglion cells.

4.2.2 Proposed Approach

This section explains two developed algorithms: 1) Histological filters which are used for extracting the cellular regions and neuroil regions; 2) Pixel depth analysis of objects and centroid mapping technique to segment nuclei, cytoplasm regions and to distinguish the ganglion cells from the rest of the cellular regions.

Histological Filters

Two major histological regions can be considered for NTs: cellular regions and neuropil regions. To segment them, their constituent mosaics must be identified. Therefore, two histological filters are developed: a cellular filter to filter the constituent mosaics of cellular regions, and a neuropil filter to extract the neuropilic regions within the image.

To construct the cellular and neuropil filters, first some of the pixels of cellular and neuropilic regions in the 25 training histological images are manually labeled as either cellular or neuropil. Then, the intensity values of all pixels with the cellular labels are stored in matrix $\mathbf{C}$ of size $c \times 3$, and all pixels with the neuropilic labels are stored in matrix $\mathbf{N}$, of size $t \times 3$, where c and t represent the total number of labeled pixels within the cellular and neuropil regions respectively, and L*, u* and v* are the three columns of each of the matrices. Both matrices are normally distributed. Then, the mean intensity value of $\mathbf{C}$ is calculated by $\bar{\mathbf{C}}_{11} = \frac{1}{c}\sum_{f=1}^{c}\mathbf{C}_{f1}$, $\bar{\mathbf{C}}_{12} = \frac{1}{c}\sum_{f=1}^{c}\mathbf{C}_{f2}$, and $\bar{\mathbf{C}}_{13} = \frac{1}{c}\sum_{f=1}^{c}\mathbf{C}_{f3}$, and the mean intensity value for $\mathbf{N}$ is computed by $\bar{\mathbf{N}}_{11} = \frac{1}{t}\sum_{d=1}^{t}\mathbf{N}_{d1}$, $\bar{\mathbf{N}}_{12} = \frac{1}{t}\sum_{d=1}^{t}\mathbf{N}_{d2}$, and $\bar{\mathbf{N}}_{13} = \frac{1}{t}\sum_{d=1}^{t}\mathbf{N}_{d3}$.

To detect the mosaics of the regions of interest, a similarity measure technique is proposed. It means that, the algorithm checks the degree of similarity between a mosaic and each of the filters. To find the degree of similarity between the

mosaics and the filters, the Euclidean distance between them is computed as follows

$$D_a(\bar{\mathbf{M}}, \bar{\mathbf{C}}) = \sqrt[2]{(\bar{\mathbf{M}}_{a1} - \bar{\mathbf{C}}_{11})^2 + (\bar{\mathbf{M}}_{a2} - \bar{\mathbf{C}}_{12})^2 + (\bar{\mathbf{M}}_{a3} - \bar{\mathbf{C}}_{13})^2} \tag{4.12}$$

and the degree of similarity between the mosaics and the neuropil filter is obtained by

$$D_a(\bar{\mathbf{M}}, \bar{\mathbf{N}}) = \sqrt[2]{(\bar{\mathbf{M}}_{a1} - \bar{\mathbf{N}}_{11})^2 + (\bar{\mathbf{M}}_{a2} - \bar{\mathbf{N}}_{12})^2 + (\bar{\mathbf{M}}_{a3} - \bar{\mathbf{N}}_{13})^2} \tag{4.13}$$

where D_a is the Euclidean distance between the a^{th} mosaic and the cellular filter or neuropil filter. Lastly mosaic i passes cellular filter or neuropil filter if $0 \leq D_a(\bar{\mathbf{M}}, \bar{\mathbf{C}}) \leq AG_C$ or $0 \leq D_a(\bar{\mathbf{M}}, \bar{\mathbf{N}}) \leq AG_N$ respectively. AG_C and AG_N are Absorption Grades for cellular and neuropil filters respectively, and they are pre-set thresholds which can be determined through prior knowledge of relevant training programs. Twenty five histological images were used from training set to tune this parameter using the procedure in section 3.3.

Neuropil Filter for Segmenting Neuropilic Regions: To extract the neuropil region, the neuropil filter is applied directly to the original image using the following expression,

$$\rho = \begin{cases} 1, & D_a(\bar{\mathbf{M}}, \bar{\mathbf{N}}) < AG_N \\ 0, & \text{Otherwise} \end{cases} \tag{4.14}$$

where ρ is a vector of length B, the total number of constructed mosaics, and ρ indicates a segmented neuropil mosaic by 1 and an eliminated mosaic by 0. The parameter $AG_N = 45$ is set through the approach in Section 3.3 by tuning on 25 training images. Figure 4.8 illustrates the output of the system after applying the neuropil filter to H&E stained histological images.

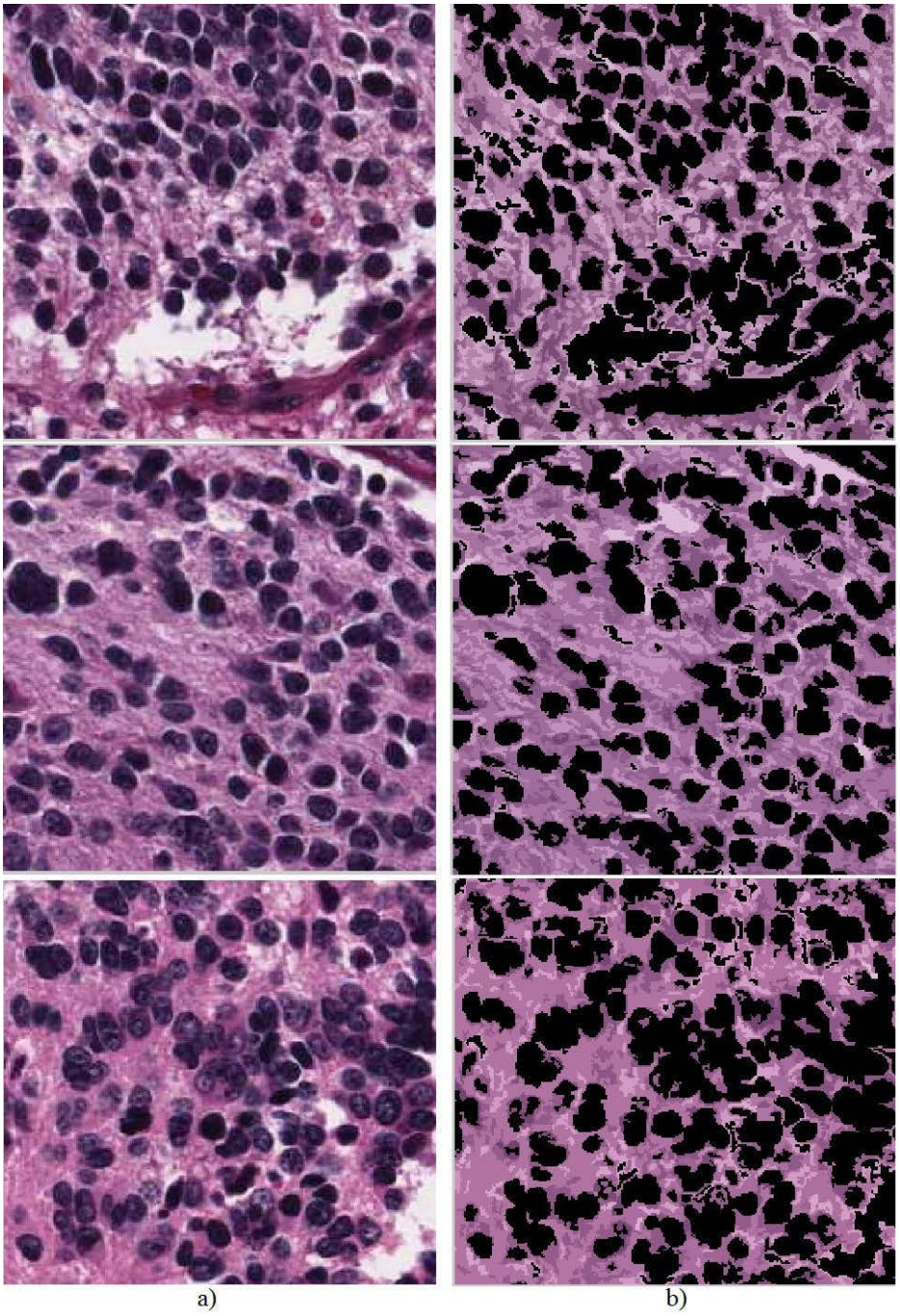


Figure 4.8: a) Three different H&E histological images, b) the obtained neuropil regions after applying the neuropil filter to the histological images.

Cellular Filter for Segmenting Cellular Regions: Cellular regions are regions distributed along the tissue and they are composed of two major compo-

nents, nuclei and cytoplasm. To extract the components, the constructed cellular filter is applied to the images using the following expression,

$$\vartheta = \begin{cases} 1, & D_a(\bar{\mathbf{M}}, \bar{\mathbf{C}}) < AG_C \\ 0, & \text{Otherwise} \end{cases} \tag{4.15}$$

where ϑ is a column vector of size B, the total number of constructed mosaics, and each element of ϑ indicates a segmented cellular mosaic by 1 and an eliminated mosaic by 0. Parameter $AG_C = 45$ is set through the approach in Section 3.3 by tuning on 25 training images. At the end of this section, it is demonstrated that the chosen value is the optimum value for AG_C to extract cellular regions. The chosen value for AG_C filters all of the components of the cellular regions, as shown in Figure 4.9. This figure illustrates that almost all of the constituent mosaics of the cellular regions pass the filter, but that some mosaics of the neuropilic regions are also permitted to pass the filter. The reason is that the intensity range of some cytoplasmic regions is close to the intensity range of the neuropil regions. Choosing a low AG_C provides a noiseless filtered image, but it eliminates many of the essential mosaics such as cytoplasmic regions. On the other hand, a filter with high AG_C allows wider intensity ranges to pass the filter, and consequently more mosaics can pass the filter, many of which are irrelevant and not part of the regions of interest. The tuned AG_C has the optimum value for segmenting all of the regions of interest with a minimum number of artifacts.

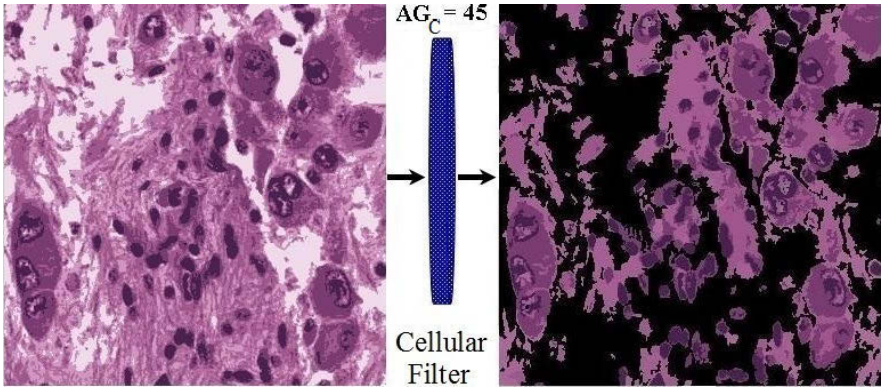


Figure 4.9: The image on the right shows the effect of the cellular filter on the enhanced histological image.

To refine the output of the obtained cellular regions and to remove the mistakenly segmented neuropil regions, the neuropil filter is applied to the cellular regions using the following expression,

$$\zeta = \begin{cases} 1, & D_a(\bar{\mathbf{M}}, \bar{\mathbf{N}}) < AG_N \text{ and } \vartheta = 1 \\ 0, & \text{Otherwise} \end{cases} \tag{4.16}$$

AG_N is the absorption grade for neuropil filter and is used to refine the cellular region by the new value of $AG_N = 15$ which is set through the approach in Section 3.3 by tuning on 25 training images and ζ is a vector of length B, the total number of mosaics within the image, and a mosaic a is part of a neuropil region within a cellular region if it has value of 1, otherwise it has value of 0. Finally, cellular regions are filtered by $\vartheta' = \vartheta - \zeta$. The output of the histological filtering is a foreground/background cellular image in which the foreground has those mosaics with label 1, and the background has a value of 0 as shown in Figure 4.10. The obtained image is of the segmented cellular regions which only contain a mixture of ganglion cells, neuroblast cells and cytoplasm regions.

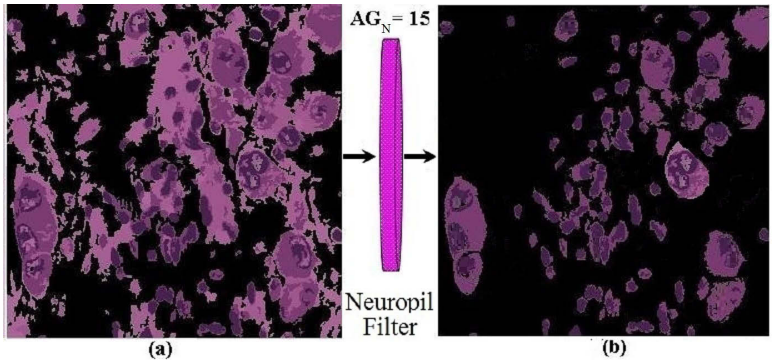


Figure 4.10: a) Unrefined cellular images, (b) refined cellular filtered images after applying neuropil filter to the unrefined image.

Segmenting Histological Structures

As mentioned in Section 2.1.1 in the literature review the main histological structures of the NTs are nuceli, cytoplasm, ganglion and neuroblast cells. This section proposes algorithms for extracting nuclei, cytoplasm and ganglion cells.

Nuclei Detection: As explained in Section 2.1.1 regarding the morphology of the ganglion cells and neuroblast cells, both types of cells contains nucleic regions. To segment the nucleic regions, the constituent mosaics of nucleic regions must be extracted from the cellular regions. Nucleic regions are darker in comparison to other components of the cellular regions, and in cellular filter lower values for AG_C filter the darker mosaics. Thus to segment the nuclei components as the darkest regions of the cellular regions, the cellular filter with a low AG_C is applied to the segmented cellular image. Consequently all the constituent mosaics of cytoplasmic regions remain behind the filter and the mosaics of nuclei pass the filter only as follows:

$$\beta = \begin{cases} 1, & D_a(\bar{\mathbf{M}}, \bar{\mathbf{C}}) < AG_C \text{ and } \vartheta'_a = 1 \\ 0, & \text{Otherwise} \end{cases} \tag{4.17}$$

where $AG_C = 15$ is obtained through the approach described in Sub-Section **??** by applying to the 25 training images and at the end of this section, it is demonstrated that the chosen value is the optimum value for AG_C. The parameter β is a column vector of length B, the total number of constructed mosaics, and each element of β indicates a segmented nuclei mosaic by 1 and an eliminated mosaic by 0. Thus, in the nucleic images the foregrounds are the constituent mosaics of the nucleic regions with label 1, and the background is 0. To reduce computational expenses the image is transformed into a binary foreground/background where all the pixels in the foreground have value 1 and the background have 0.

Cytoplasm Detection: All the cells contain nuclei but not all of them contain cytoplasmic components. Discernible cytoplasm around the nuclei is indicative of cell maturation and indiscernible cytoplasm is indicative of immature cells.

To segment the cytoplasmic regions, the system must analyze the regions around the nuclei. The centroid mapping technique is proposed because it determines the differences between the surrounding of nucleic regions in the nucleic image and cellular images. Figure 4.11 gives an example of nucleic and cellular images. As shown in the figure, for each nucleus in the nucleic image there is a nucleus in the cellular image. However, in the cellular image some of the nuclei are surrounded by cytoplasmic components. For a nucleus with cytoplasmic components, the distance from its centroid to the background in the nucleic image is not equal to the distance from its centroid to the background in the cellular image.

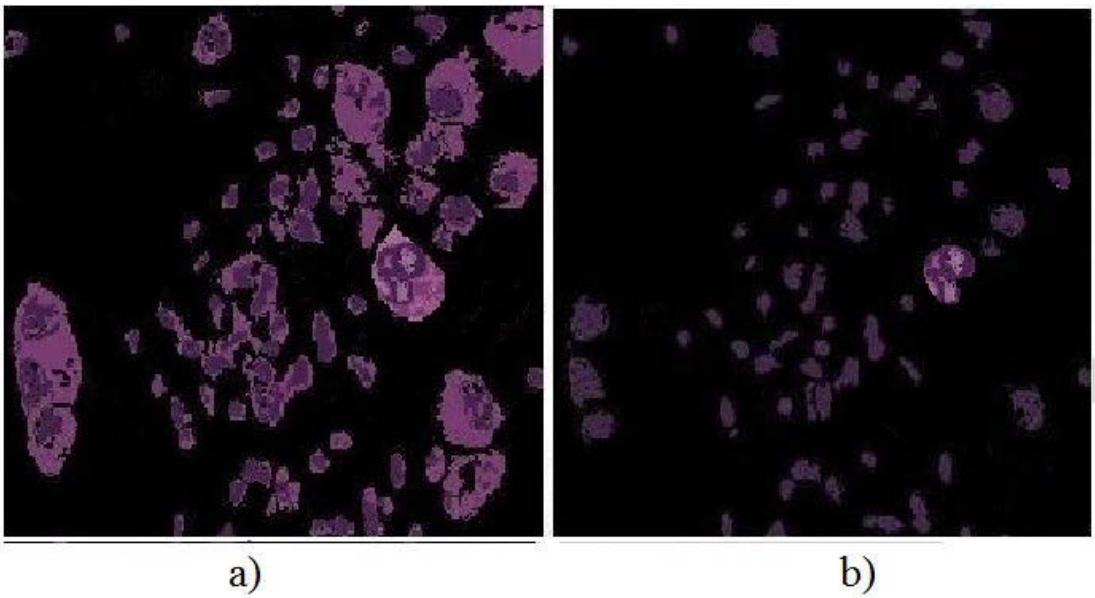


Figure 4.11: a) The obtained cellular image after applying $AG_C = 45$ and $AG_N = 15$, (b) nucleic image after applying $AG_C = 15$ to the cellular image.

To find the centroid, the distance between each of the constituent pixels of nucleic regions and the background of the image is calculated. The pixels with greatest distance to the background are chosen as the centroid. To achieve this, the Euclidean distance of each of the constituent pixels of the nucleic regions from the nearest background is calculated; an example is shown in Figure 4.12a. Those pixels with the maximum distance from the background are in the centroid of the nucleus. The next step is to compute the Euclidean distance between the pixels on the centroids of the nuclei to the backgrounds through the eight neighbor pixels, as shown by the arrows in Figure 4.12b. The distances between constituent pixels of each centroid to the background on the nucleic image through the eight neighbor pixels are stored in centroid nuclei matrix $\mathbf{W}_j$ for $j \in \{1, ..., S\}$. The size of the matrix is $b \times 8$, and S and b are the total number of nucleic centroids in the nucleic image and the total number of pixels in the nucleic centroid j respectively. A centroid can be constructed from more than one pixel. Each row

in $\mathbf{W}_j$ indicates a constituent pixel of centroid j, and eight columns for each row represents the distance of the pixel to the background through the eight neighboring pixels in the nucleic image.

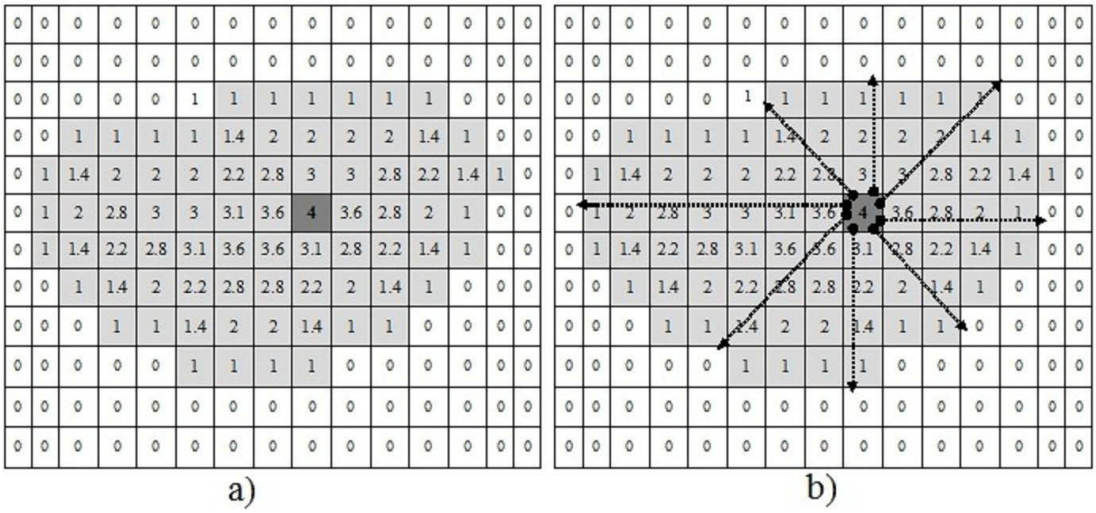


Figure 4.12: a) The Euclidean distance between each pixel and the nearest background (0), b) The Euclidean distance between the centroid and the background through the eight neighbor pixels.

The next step is to determine the presence of cytoplasmic regions around the nuclei. To this end, first the cellular image is converted into binary format. Then, the algorithm computes the distances of nuclei centroids to background in both nuclei image and cellular image. If the distance from the centroids of nuclei to backgrounds in the nuclei image is not equal to the distance from the same centroids of nuclei to the backgrounds in the cellular image, it means that there is a cytoplasmic region around the nuclei. Then the pixels on the centroids of the nuclei in $\mathbf{W}_j$ are mapped into the cellular image to find the centroid of the nuclei of the cellular image. To compute the distance between the centroids of the nuclei and backgrounds in the cellular image, the Euclidean distance from the centroids to the backgrounds is computed through the eight neighbor pixels and the results are stored in matrix $\mathbf{Q}_j$ of size $b \times 8$ where b is the total number of

pixels in centroid j, and the eight columns indicate the distances from the eight neighbor pixels to the background. Using $\mathbf{\Delta}_j = \mathbf{Q}_j - \mathbf{W}_j$, the distance variation between the centroids of the nuclei to the background in the nucleic image and the centroids of the nuclei to the background in the cellular image are obtained. Lastly, the average distance of eight neighbor pixels for all pixels at the centroid j in $\mathbf{\Delta}_j$ is computed and stored in $\mathbf{av}_j$ with size $b \times 1$. Then, the average distance of all b pixels of centroid j in $\mathbf{av}_j$ are stored in $\bar{\mathbf{av}}_j$ for $j \in \{1, ..., S\}$. To extract the ganglion cells the following expression is used,

$$\varrho = \begin{cases} \text{Discernible Cytoplasm}, & \bar{\mathbf{av}}_j \geq \lambda, \\ \text{Indiscernible Cytoplasm}, & \text{Otherwise} \end{cases} \tag{4.18}$$

and $\lambda = 5$ is a preset threshold which was obtained through the approach in Section 3.3 by tuning on the 25 training images. During training of the algorithm different values for λ were taken into consideration and $\lambda = 5$ provide the optimum performance in recognizing overlapping cells. This means a cytoplasm is considered as discernible by the pathologist when it has at least a diameter of 5 pixels.

Segmenting Ganglion Cells: Determining cytoplasmic regions around the nuclei of the cells facilitates the process of segmentation for ganglion cells. The main morphological property of a ganglion cell is the large amount of cytoplasm around the nucleus. According to Shimada et al. (2001) for a ganglion cell, the size of the cell is twice the size of the nuclei. Thus, to find the ganglion cells, the system must analyze the size of identified cytoplasmic regions around the nuclei of each of the cells. To this end, the average distance of every pixel at centroid j to the background in $\mathbf{W}_j$ and $\mathbf{Q}_j$ are computed and stored in $\mathbf{aw}_j$ and $\mathbf{aq}_j$, each

vectors of length b. Then, the average distance of all b pixels of centroid j in $\mathbf{aw_j}$ and $\mathbf{aq_j}$ is stored in $\mathbf{\bar{aw}_j}$ and $\mathbf{\bar{aq}_j}$ respectively. Finally to extract the ganglion cells the following expression is used,

$$\omega = \begin{cases} \text{Ganglion}, & \frac{\mathbf{\bar{aw}}_j}{\mathbf{\bar{aq}}_j} \geq 0.5, \\ \text{Other Cellular Components}, & \text{Otherwise} \end{cases} \tag{4.19}$$

The outputs of the proposed algorithm are shown in Figure 4.13.

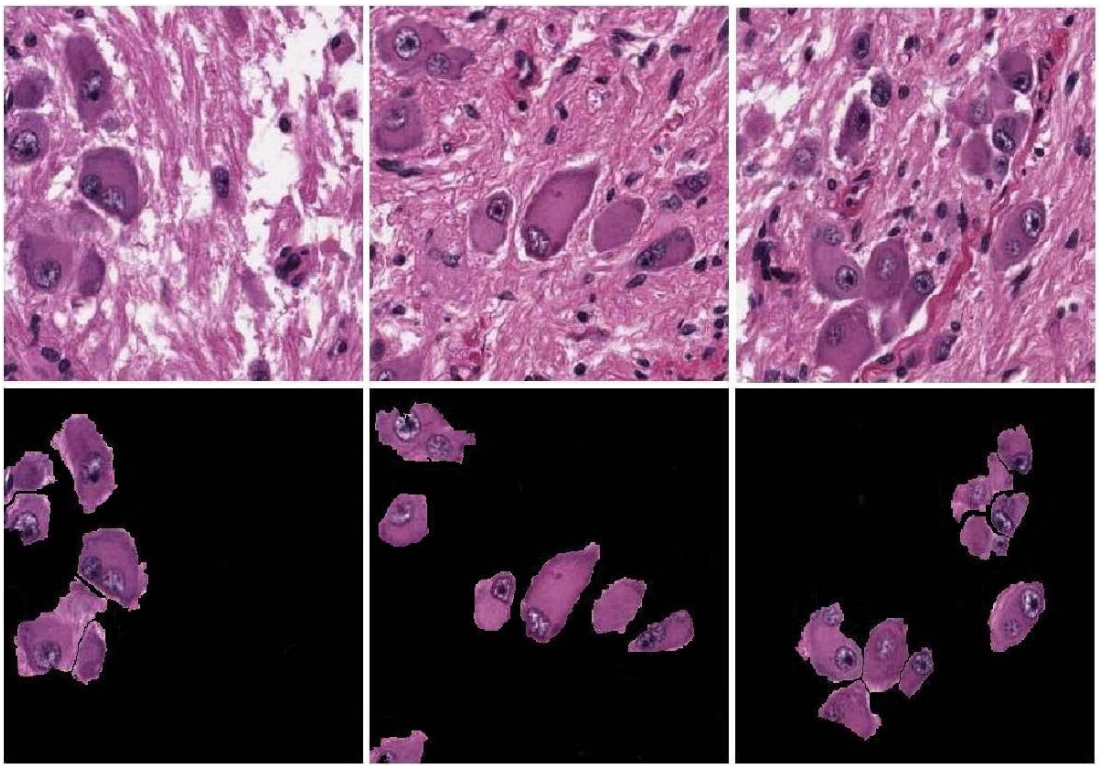


Figure 4.13: The top three images are three different histological images which contain different types of cells, and the three images at the bottom are the extracted ganglion cells after applying the segmentation approach to the three histological images.

The proposed algorithm contains a set of parameters (such as Absorption Grades (AG)) which must be tuned for segmenting cellular regions. To show that the chosen AGs are tuned to the optimum values, different values for AGs

are tested on 25 training images and the result are shown in Figure 4.14. The figure indicates different F-measures of the system in segmenting ganglion cells for different values of AGs. The figure shows that the best performance is obtained by a combination of AG_C=45 and AG_N=15 for segmenting cellular regions, and AG_C=15 for segmenting nucleic regions for determining ganglion cells.

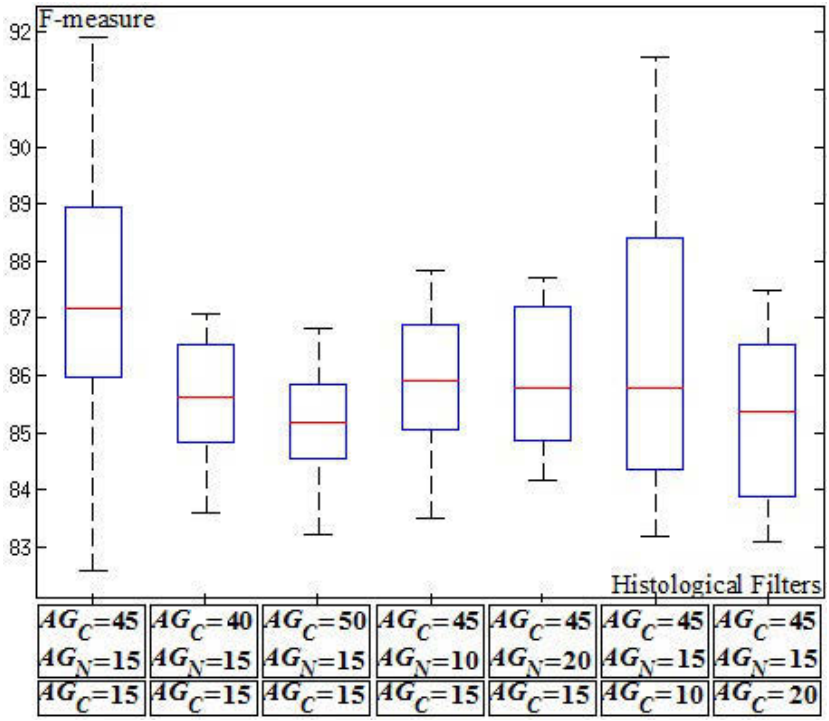


Figure 4.14: Box plot of F-measure over 25 training images for different Absorption Grades (AG)s values. AGs that are used in the developed histological filters to filter the cellular regions and the nucleic component. The first two AGs in a rectangle indicate the combination of AGs for segmenting cellular regions and AGs in the other rectangle are used for segmenting the nucleic component.

4.2.3 System Validation and Experimental Results for Segmenting Cellular Regions and Ganglion Cells

All ganglion cells in both the training set and the test set were manually segmented by an independent pathologist. Finally, to validate the developed sys-

Table 4.1: Average of True Positive/Pathologist results of segmenting ganglion cells, False Positive, False Negative, precision, recall and F-measure of the proposed algorithms in this thesis.

	TP / Path	FP	FN	P	R	F-measure
Ave	38.32 / 42.24	6.44	4.32	85.57%	89.55%	87.47%
StDve				2.74	3.27	2.32

tem, the results of the system are compared with those of the pathologist. Then Precision, Recall and F-measure are used to measure the accuracy of the system in segmenting ganglion cells.

Experiments were run to compare the performance of the system in segmenting cellular regions and ganglion cells with four other systems which are described in Section 2.2: 1) the system proposed by Lu et al. (2013), 2) the system developed by Nam et al. (2012), 3) the system introduced by Sadeghian et al. (2009), and 4) the system proposed by Ta et al. (2009). They are applied to the test images for segmenting the ganglion cells.

The system is applied to the 25 histological test images, and the results of the system in counting the ganglion cells are compared with those of the pathologist. Table 4.1 indicates the average performance of the developed system in segmenting ganglion cells by F-measure, precision, recall, number of True Positive (TP), False Positive (FP) and False Negative (FN). The average number of ganglion cells per image obtained by the pathologist is 42.24, and the system counted 38.32 ganglion cells correctly on average per image. The average FP, that is the average number of ganglion cells identified by the developed system but rejected by the pathologist is 6.44 cells per image. The average FN of the system, that is the average number of cells not identified by the developed system, is 4.32 ganglion cell per image.

Table 4.2: Results of the proposed systems in the state of the art in segmenting ganglion cells.

Systems	TP	FP	FN	P	R	F-measure	StdDev
The proposed System	38.32	6.44	4.32	85.57%	89.55%	87.47%	2.32
Lu et al.	35.88	7.09	6.01	83.50%	85.65%	84.56%	2.38
Nam et al.	35.39	7.56	5.9	82.39%	85.71%	84.02%	2.42
Sadeghian et al.	32.81	8.11	7.59	80.18%	81.21%	80.69%	3.39
Ta et al.	29.88	8.49	7.79	77.87%	79.32%	78.59%	2.57

Table 4.2 indicates the performance of state of the art method for segmenting ganglion cells after their application to the test images. Among those systems, the best average F-measure is 84.56% obtained by Lu et al. (2013). This is approximately 3% less than the results obtained by the method proposed in this thesis. Ta et al. (2009) provides the lowest overall F-measure of 78.59%. Ta et al. (2009) method shows the lowest precision which means the highest over-segmentation, whereas Lu et al. (2013) obtained the highest precision which is 2% lower than the precision obtained by the developed system in this thesis. The lowest average recall is also generated by Ta et al. (2009). Table 4.2 also indicates that after the system proposed in this thesis, the system proposed by Lu et al. (2013) has the lowest average standard deviation for F-measure among all other systems. As explained in the literature review the method proposed by Nam et al. (2012) generates the highest standard deviation.

Also to show that the results for segmenting ganglion cells obtained by the proposed algorithms in this thesis are significantly better than the results obtained by the state of the art algorithms, Holm statistical hypothesis test is applied (Holm, 1979). Table 4.3 illustrates the results of Holms statistical tests which

were applied to indicate that the algorithm proposed in this thesis significantly outperforms the above state of the art algorithms in segmenting ganglion cells. The null hypothesis of the tests is defined as similarity between precision, recall and F-measure of the algorithm proposed in this thesis and those of the algorithm proposed by Lu et al. (2013). The algorithm proposed by Lu et al. (2013) is chosen for the statistical tests among the others, because it obtained the best performance among other state of the art algorithms in Table 4.2. Table 4.3 shows that p-values for precision, recall and F-measure of the tests are far less than the level of significance, and consequently the hypothesis is rejected. This concludes that the algorithm proposed in this thesis significantly outperforms the algorithm proposed by Lu et al. (2013), and the other algorithms in the experiments.

The above stated algorithms as described in Section 2.2 are based on morphological, active contour and thresholding techniques. The results indicate that the main issue with the algorithms based on morphological techniques such as the algorithm in Lu et al. (2013) and the thresholding technique such as the algorithm in Ta et al. (2009) is sensitivity to the quality of images. The issue with algorithms based on the active contour technique in Nam et al. (2012) and Sadeghian et al. (2009) is sensitive to noise on the contour of the cells. The results of the algorithm proposed in this thesis show that it significantly addresses the issues with the above systems.

Table 4.3: Holms statistical tests for precision, recall and F-measure of the proposed system in this thesis and Lu et al. for segmenting ganglion cells with $\alpha = 0.1$.

Ganglion Cell Segmentation	z_0	p-value
Precision	1.799	0.071
Recall	4.2	2.67×10^{-6}
F-measure	3.8	1.44×10^{-5}

4.2.4 System Validation and Experimental Results for Segmenting Neuropil Regions

To evaluate the performance of the proposed system in computing the amount of neuropil regions within each of the tumor, a simple software application was developed for this thesis to allow the pathologist to mark the neuopil regions manually as shown in Figure 4.15. The software stores the x and y coordinates of all the segmented pixels in a list called list of manually segmented pixels. If a region is mistakenly marked more than once by the pathologist, the constituent pixels of that region are only counted once by the software. To evaluate the performance of the proposed algorithms, they are applied to the test images and x and y coordinates of all the segmented pixels are stored in a list called the list of automated segmented pixels. If the x and y coordinates of a pixel in the list of automatically segmented pixels has a corresponding x and y coordinates in the list of manually segmented pixels, then it is considered as a True Positive (TP). If the coordinates of a pixel in the list of automatically segmented pixels do not have any corresponding coordinates in the list of manually segmented pixels, then the pixel is considered as a False Positive (FP). Finally, if the coordinates of a

pixel are listed in the list of manually segmented pixels, but it do not have any corresponding coordinates in the list of automatically segmented pixels, then the pixel is considered as a False Negative (FN). Also the results of the proposed system are compared with three other robust and well known segmentation methods reviewed in Section 2.2.1, Russell et al. (2009), Kittler and Illingworth (1986), and Ihlow et al. (2011). Although the proposed methods by Russell et al. (2009) and Ihlow et al. (2011) are developed to segment the cellular regions, their algorithms can be adopted to detect and extract neuropil regions as well. Because both of the algorithms analyze the intensity values of pixels to detect different types of histological regions, they can be used to find the pixels which show the intensity level of the neuropil regions.

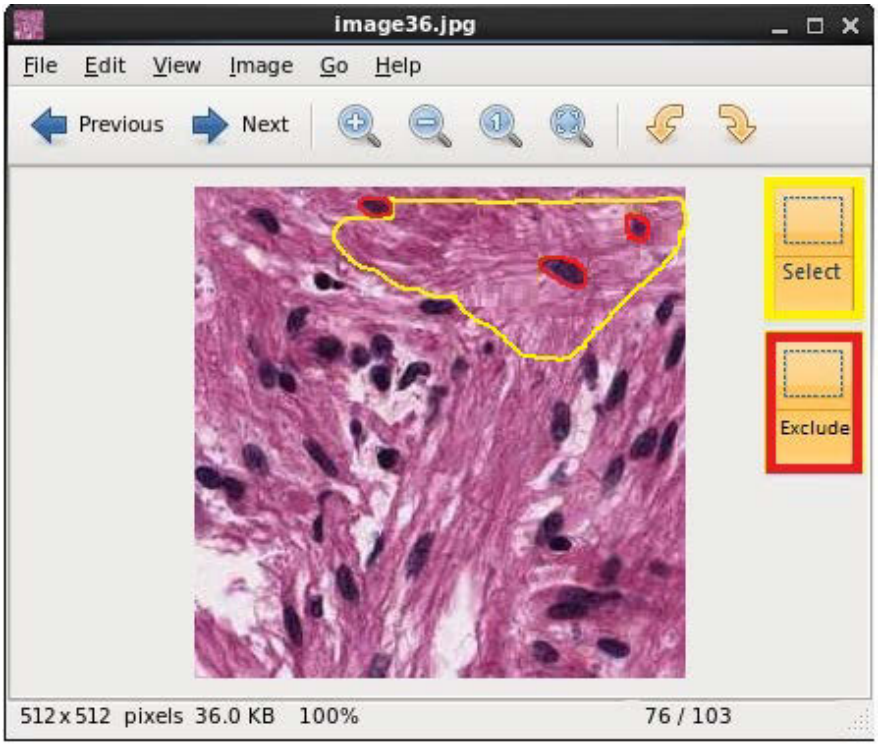


Figure 4.15: Interface of the developed software application for pathologist to manually segment the neuropil regions.

Table 4.4 indicates the average precision, recall and F-measure of the system

proposed in this thesis and the state of the art systems in measuring the amount of neuropil within the test 25 images. The results indicate that the proposed method in this thesis achieves superior performance in precision, recall and F-measure in comparison with all other state of the art methods, while the method proposed by Ihlow et al. (2011) obtains the lowest performance among the other methods in measuring the amount of neuropil within the image of NTs. Moreover, the standard deviation obtained by the proposed system in this thesis is less than that of all the other state of art systems which indicates the consistency of the system in segmenting neuropil regions in different histological images.

Table 4.4: Average performance of the system proposed in this thesis and state of the art systems in measuring the amount of neuropil.

Systems	Precision (Avg)	Recall (Avg)	F-measure (Avg)	SetDev
The proposed system	91.51%	96.65%	94.01%	2.09
Russell et al. (2009)	86.81%	88.43%	87.61%	2.17
Thresholding (Kittler)	85.99%	75.47%	80.38%	2.31
Ihlow et al. (2011)	82.71%	73.33%	77.73%	2.11

Table 4.5 illustrates the results of Holms statistical tests which were applied to indicate that the algorithm proposed in this thesis significantly outperforms the above state of the art algorithms in segmenting neuroblast cells. Table 4.5 shows that p-values for precision, recall and F-measure of the tests are far less than the level of significance, and consequently the hypothesis is rejected. This concludes that the algorithm proposed in this thesis significantly outperforms the algorithm proposed by Russell et al. (2009), and the other algorithms in the experiments.

Table 4.5: Holms statistical tests for precision, recall and F-measure of the proposed system in this thesis and Russelli et al. for segmenting neuropil regions with $\alpha = 0.1$.

Neuropil Segmentation	z_0	p-value
Precision	5.208	1.903×10^{-7}
Recall	3.59	3.280×10^{-4}
F-measure	4.849	1.239×10^{-6}

4.3 Segmenting Neuroblast Cells

One of the major prognostic factors for NTs is the total number of neuroblastic cells within the tumor. The ratio of ganglion cells to neuroblast cells is an indicator of distinguishing poorly differentiated from differentiated NTs. Neuroblastic cells have blue nuclei and indiscernible cytoplasm, and they are parts of the cellular regions. The total number of neuroblastic cells within the tumor depends on the maturation of the tumor. Immature tumors have an enormous number of neuroblastic cells, while more mature tumors have fewer neuroblastic cells. Figure 4.16 illustrates an H&E stained histological slide of NT that contains neuroblastic cells. This section continues to address contribution 2 by segmenting and counting the total number of neuroblastic cells within the images derived from H&E stained histological slides. To segment the neuroblast cells, the system must take three stages into the consideration. 1) Extracting cellular regions from the image. 2) Distinguishing the neuroblast cells from other cellular components in the image. 3) Segmenting neuroblastic cells from the cellular regions.

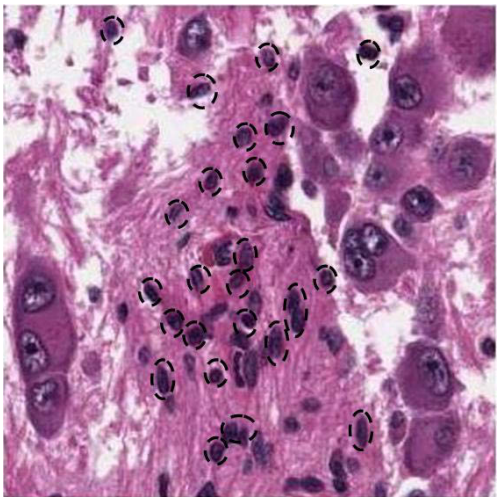


Figure 4.16: Dashed circles indicate neuroblastic cells with round blue nuclei and indiscernible cytoplasm in H&E stained tissue slides.

4.3.1 Methodology

To address this issue, a hybrid algorithm is proposed which integrates the developed pre-segmentation technique, the proposed histological filters, the centroid mapping technique and the adaptive thresholding technique. After applying the pre-segmentation technique, the cellular regions are segmented by the adaptive thresholding based on Otsu method (Otsu, 1975). Ganglion cells are segmented from the cellular image using the proposed histological filter and centroid technique. Finally, by applying the proposed cellular filters, the system extracts the neuroblast cells from the rest of the histological structures such as the cytoplasmic structures within the cellular regions. Figure 4.17 indicates the general view of the proposed algorithm for segmenting neuroblast cells.

To test and train the system, 10 different tissue arrays were used. To construct the test set, 20 tissue spots were taken from 5 tissue arrays randomly. To build the training set, 20 tissue spots were taken from the other 5 tissue arrays that were different from the test set.

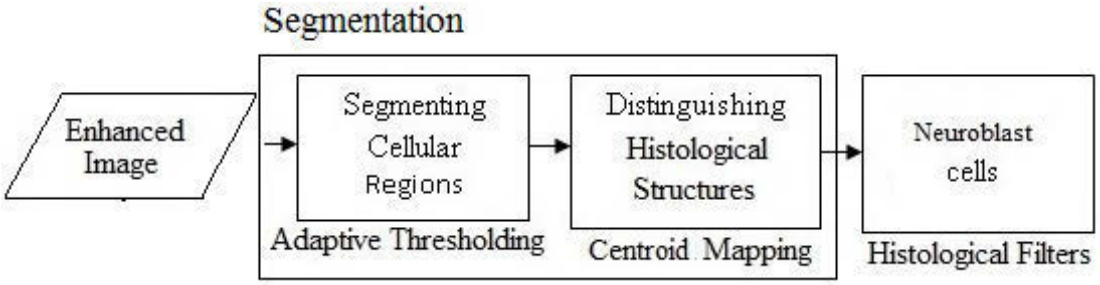


Figure 4.17: General view of the proposed algorithm for segmenting neuroblast cells.

4.3.2 Proposed Approach

Three stages are considered for segmenting the neuroblast cells as one of the histological structures of cellular region. The first step is to segment the cellular regions. The second step is to distinguish between ganglion cells and neuroblast cells. The third step is to extract the neuroblast cells.

Segmenting Cellular Regions

As mentioned in Section 2.1, blue regions within the tumor tissue represent cellular regions, and this characteristic of cellular regions is used to segment them. It means that the specific intensity level of cellular regions is the basis for cellular segmentation. The first step in segmentation is to apply the proposed pre-segmentation algorithms introduced in Section 4.1 of this chapter to the images. Then the proposed histological filters in Section 4.2.2 of this chapter are applied to the enhanced images. The histological filters extract the constituent mosaics of cellular regions. Figure 4.18 indicates the output of the system after applying the cellular filter.

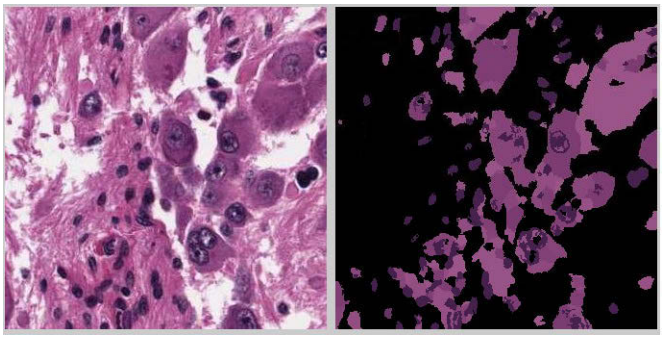


Figure 4.18: Segmented cellular regions which includes nuclei and cytoplasm regions using the developed cellular filter and neuropil filter.

The output is a foreground/background image where the foreground shows cellular regions and background is 0. The segmented cellular regions contain two types of cells (ganglion cells and neuroblast cells) and cytoplasm components.

Distinguishing Different Types of Histological Structures

To distinguish the ganglion cells from rest of the histological structures within the cellular regions, the proposed centroid mapping technique which is introduced in cytoplasm detection of Section 4.2.2 is applied to the segmented cellular images. This technique locates the ganglion cells by analyzing the cytoplasm regions around the nuclei in the image. A cell that contains a nucleus which is surrounded by cytoplasmic component is considered as a ganglion cell. Then all of the identified ganglion cells are eliminated from the cellular regions. Figure 4.19 indicates the output of the system after eliminating the ganglion cells. As a result, the remaining segmented cellular regions contain neuroblast cells and some cytoplasmic components which do not belong to any nucleic components. Cytoplasmic components must usually be accompanied by nucleic components, but during the process of biopsy some of the them are parted from their nucleic components as shown by circles in the figure.

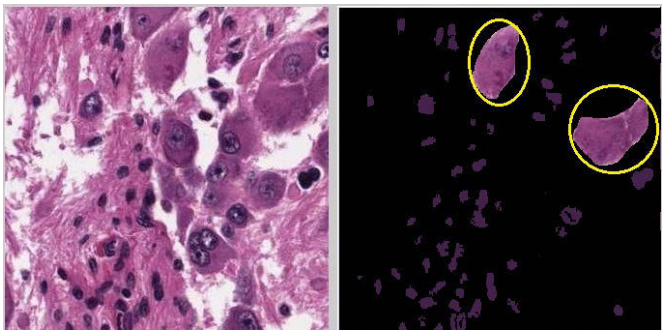


Figure 4.19: The left image is the original image and the right image is the segmented cellular regions which includes nuclei and cytoplasm regions which are marked by the circles.

Extracting Neuroblast Cells

Finally to segment the neuroblast cells and distinguish them from the cytoplasmic components, adaptive thresholding based on the Otsu method is applied to the segmented cellular images. As explained in Section 2.2 of this thesis, the Otsu method efficiently and accurately determines a different range of intensity values using the grey-level histogram of the pixels in the images with low range of intensity variation. The developed pre-segmentation technique reduces the intensity variations between the pixels in the images, therefore the Otsu method provides high performance in terms of segmenting the regions of interest. It binarises the image into foreground and background. The foreground consists of pixels whose intensity values are greater than the determined threshold, and the background consists of those pixels that fail to pass the threshold and whose values are zero.

As explained in Section 4.1 the color spaces of the images are transformed from RGB to L*u*v* during the process of pre-segmentation. The Otsu method is applied to luminance coordinate (L*) of the L*u*v* only to find the optimum thresholding value of the cellular regions. The reason for this is that the Otsu method only analyses the gray-level histogram of the pixels, and two other co-

ordinates provide chrominance information. Figure 4.20 illustrates the extracted neuroblast cells after applying thresholding technique.

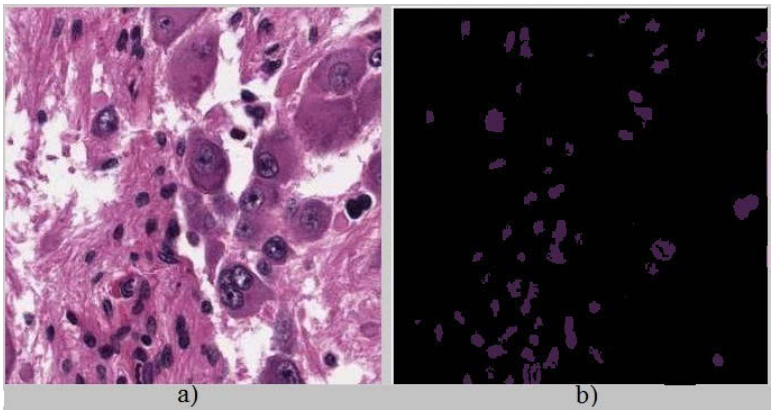


Figure 4.20: a) Original image, b) segmented neuroblast cells.

4.3.3 System Validation

The discrepancy between the estimated number of neuroblastic cells using the proposed system and those of pathologist, shows the accuracy of the system. A correctly segmented cell is considered to be a True Positive (TP), a cell detected by the system but rejected by the pathologist to be a False Positive (FP), and a cell that is not detected is considered to be a False Negative (FN). There is no True Negative (TN) because neuroblastic cells are the only regions of interest in this research. Using the above-stated parameters, the precision, recall and F-measure (Powers, 2011) of the system are calculated.

The proposed method in this thesis is compared with three other state of the art methods in terms of segmenting the neuroblast cells: the method proposed by Yang et al. (2006), the method proposed by Begelman et al. (2004), and 3) adaptive thresholding using the Kittler method (Kittler and Illingworth, 1986). The details of these systems were described in Section 2.2.

Table 4.6: The obtained Precision, Recall and F-measure on neuroblast cell counting for the developed system in this thesis in comparison with those of the pathologist.

Image	Precision	Recall	F-measure
1	85.87%	82.19%	83.98%
2	84.76%	81.81%	83.25%
3	85.81%	88.57%	87.16%
4	87.61%	87.06%	87.33%
5	84.63%	80.56%	82.54%
6	84.76%	88.78%	86.72%
7	82.74%	86.11%	84.39%
8	81.96%	80.15%	81.05%
9	85.04%	92.10%	88.42%
10	83.21%	91.53%	87.17%
11	88.16%	85.65%	86.88%
12	85.03%	88.76%	86.85%
13	83.61%	84.16%	83.88%
14	83.73%	84.50%	84.11%
15	84.37%	83.86%	84.11%
16	85.87%	83.78%	84.81%
17	84.49%	86.15%	85.31%
18	85.77%	82.48%	84.09%
19	87.21%	88.10%	87.65%
20	83.38%	80.54%	81.93%
Avg	**84.9%**	**85.34%**	**85.08%**
StdDev	**1.60**	**3.55**	**2.08**

4.3.4 Experimental Results

To test the developed algorithms, 20 histological images are used, and accuracy of the system in terms of counting the total number of neuroblastic cells is measured by comparing with the pathologist's results. Table 4.6 indicates the precision, recall and F-measure of the system.

Three different experiments are run to evaluate the performance of the proposed algorithm in comparison with other state of the art algorithms. Figure

4.21 indicates the precision and recall of the methods proposed in this thesis, and the state of the art methods. As the figure indicates, the precision and recall obtained by the proposed system in this thesis are greater than those of state of the art algorithms.

Figure 4.22 illustrates the F-measure of the proposed algorithm in terms of counting the total number of neuroblastic cells. It provides higher performance compared to other algorithms. The average F-measure of the proposed algorithms in this thesis is 85.08%, the average F-measure of Yang et al. is 83.85, Begelman et al. is 83.02%, and Kittler is 80.28%.

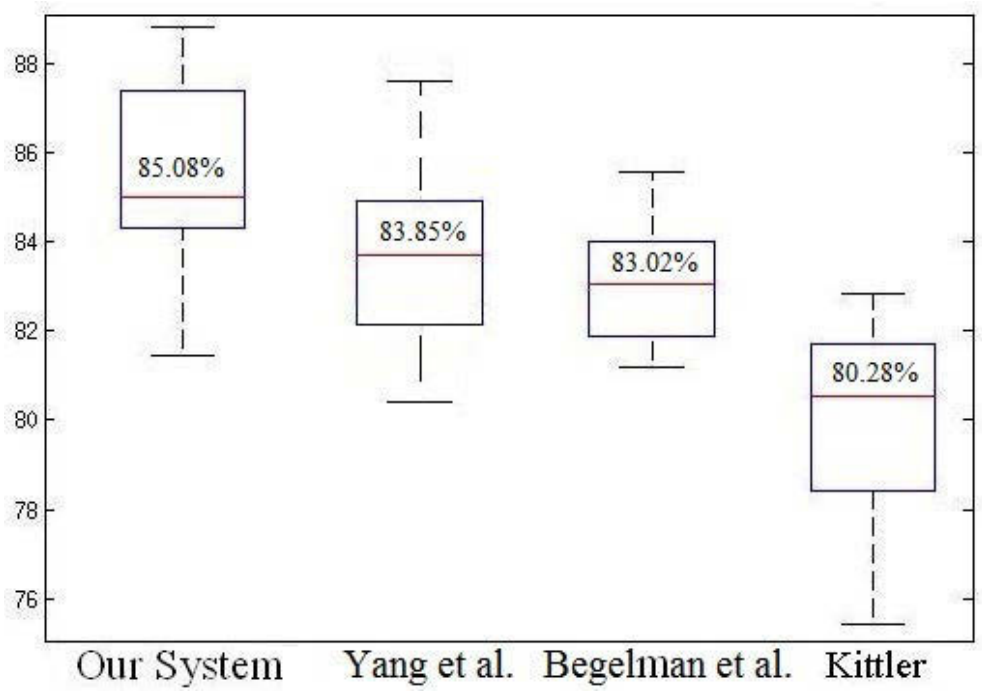


Figure 4.21: The overall F-measure of the method proposed in this thesis, Yang et al. (2006), Begelman et al. (2004) and Kittler

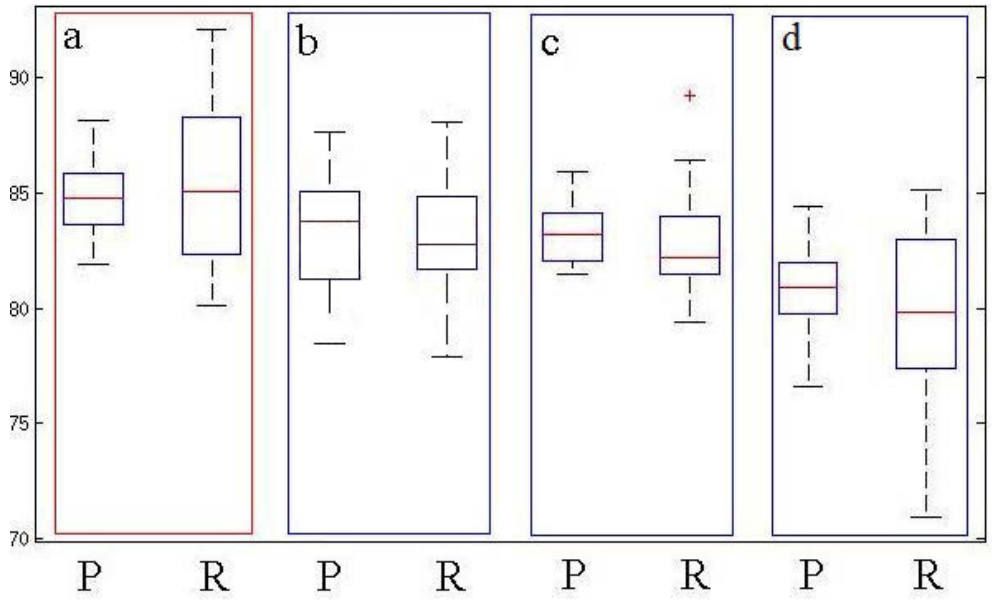


Figure 4.22: a) Precision (P) and Recall (R) of the method proposed in this thesis, b) P and R of Yang et al. (2006), c) P and R of Begelman et al. (2004), and d) P and R of Kittler

Table 4.7 illustrates the results of Holms statistical tests which were applied to indicate that the algorithm proposed in this thesis significantly outperforms the above state of the art algorithms in segmenting neuroblast cells. Table 4.7 shows that p-values for precision, recall and F-measure of the tests are far less than the level of significance, and consequently the null hypothesis is rejected. This concludes that the algorithm proposed in this thesis significantly outperforms the algorithms proposed by Yang et al. (2006), and the other algorithms in the experiments.

Table 4.7: Holms statistical tests for precision, recall and F-measure of the proposed system in this thesis and Yang et al. for segmenting neuroblast cells with $\alpha = 0.1$.

Neuroblast Cell Segmentation	z_0	p-value
Precision	1.78	0.07
Recall	2.68	0.007
F-measure	3.13	0.0001

4.4 Related Publications

All of the proposed algorithms in this chapter are published in peer-reviewed journals and conference papers. The proposed algorithms in Sections 4.1, 4.3 and 4.4 was published and submitted to peer-reviews journals and conference papers. The articles related to this chapter are listed in below,

Refereed journal papers,

- S. Tafavogh, K. Felix Navarro, D. R. Catchpoole, and P. J. Kennedy. "Non-parametric and integrated framework for segmenting and counting neuroblastic cells within neuroblastoma tumor images". *Medical and Biological Engineering and Computing*, Springer, 51, pp: 645–655, 2013.

- December 2013: submitted to journal of BMC Bioinformatics, S. Tafavogh, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy "Segmentation and quantitative analysis of ganglion cells using histological and topological filters".

Refereed conference papers,

- S. Tafavogh, D.R. Catchpoole, P.J. Kennedy. "Determining cellularity status of tumors based on histopathology using hybrid image segmentation". IEEE World Congress on Computational Intelligence, Australia, pp:1-8, 2012.

- S. Tafavogh. Comparing clustering methods and color-spaces for histological images, In Proceedings of the IEEE International Symposium on Image and Signal Processing and Analysis, pp: 154-59, 2011.

- S. Tafavogh. "Effect of HSV and L*a*b* color spaces on segmenting histological images by expectation maximisation algorithm", Conference on Computational Molecular Biology and Bioinformatics, Moscow, vol. 22, pp: 359-361, 2011.

- S. Tafavogh. "The effects of adaptive histogram equalization and color spaces on histological segmentation quality of expectation maximization method". International Conference on Graphic and Image Processing (ICGIP), October 2011, Cairo, Egypt.

4.5 Contribution, Discussion and Conclusion

Making a prognosis for cancer requires accurate microscopic assessment and quantitative analysis of cellular regions in the tumor tissue. Prognosis of neuroblastoma is highly dependent on the ratio of neuropil regions, cellular regions and the number of a specific type of nerve cells, such as ganglion cells and neuroblast cells within the NTs. In this section, a series of algorithms was proposed to locate and segment the above-stated prognostic factors.

The main problems with developing such systems are; 1) The fluctuating

quality of H&E stained histological slides, which significantly reduces the performance of the system in locating and determining different histological regions and histological structures, and 2) Distinguishing different histological regions and histological structures and extracting them from the histological images.

This chapter addresses Contribution 1 and Contribution 2 of this thesis as listed in Section 1.5 by improving the quality of the H&E stained histological images, and proposing segmentation algorithms to extract different types of histological regions and histological structures respectively.

Contribution 1, a novel global and local intensity variation reduction approach is proposed in Section 4.1 to enhance the quality of the images. Color quantization techniques are deployed to reduce the range of different intensity levels among the images. Also for local intensity variation reduction a mosaicking approach is developed which groups the constituent pixels of the histological regions based on their joint intensity-spatial domain, and reduces the intensity variation within each of the mosaics. Two advantages of the proposed pre-segmentation method are; 1) reducing the intensity variation between constituent pixels of similar histological regions. This results in increasing the segmentation accuracy. 2) Reducing the computational expenses by considering a mosaic with all of its constituent pixels as one big pixel. Thus, the system analyzes images mosaic by mosaic instead of pixel by pixel.

Contribution 2, as described in Section 4.2 develops methods to locate and extract different histological regions such as cellular regions and neuropil regions and histological structures such as cytoplasm nuclei, ganglion cells and neuroblast cells. For segmenting cellular regions and neuropil regions, two histological filters are developed as the cellular filter and neuropil filter. The cellular filter segments the constituent mosaics of the cellular regions and the neuropil filter extracts

the neuropil regions and refines the output of the cellular filter as well. Also the cellular filter is used for extracting the nucleic structures. To segment the cytoplasmic regions and ganglion cells, a centroid mapping and pixels depth analysis techniques are developed. For segmenting neuroblast cells within the NTs, a novel hybrid segmentation technique is proposed by integrating the developed pre-segmentation algorithm with the developed histological filter, the centroid mapping approach and the adaptive thresholding method. The advantages of the algorithms are insensitivity to the size and orientation variation of histological structures.

The outcomes of the above stated algorithms are compared with other robust state of the art systems. The results of the proposed algorithm are far superior to that of the other systems.

The main issue with the results of neuroblast algorithms is overlapping cells. It means that the system is not able to detect the overlapping neuroblast cells from single neuroblast cells. This is the main reason that the results obtained by the proposed algorithm do not obtain far superiority over the algorithm by Yang et al. (2006). Images of the segmented neuroblast cells contain a number of overlapping neuroblast cells which decreases the accuracy of the system in counting the total number of the cells. Consequently, in the following chapter post-segmentation techniques are developed to improve the performance of the system in counting the total number of neuroblast cells.

Chapter 5

Post-Segmentation: Splitting Overlapping Cells

The main issue with developing automated cellular quantitative analysis systems is dealing with the error imposed by overlapping cells. An enormous number of overlapping cells in the slides and the inability of the system to distinguish between different types of cells significantly reduces the accuracy of automated cell counting (Al-Kofahi et al., 2010).

This chapter addresses **Contribution 3** of this thesis by handling the issue of counting overlapping neuroblast cells in images of NTs. The proposed method in this thesis overcomes many of the limitations of the cell counting algorithms which are reviewed in the literature such as over-segmentation, the false concave regions on the contour of cells, and the high sensitivity to the quality of the histological images.

The proposed approach distinguishes between overlapping cells and single cells by the differences in their morphological properties. Then novel overlapping cell splitting algorithms are proposed. The system uses the dominant concave regions

of the cells as markers to identify the overlap region. Dominant concave regions are identified by analyzing the convex-hull and area of the cells. The system then finds the initial splitting points at the critical points of the concave regions. To do so, the obtained concave regions are decomposed into their components such as arcs, chords and edges, and the distance between the components are analyzed using a developed seed growing technique to find the initial splitting points. Finally, to determine the optimum splitting route between two candidate initial splitting points, a shortest path determination approach is developed.

The proposed algorithms in this chapter are not sensitive to the staining quality of the tissues and the false concave regions on the contour of the cells, and they reduce the over-segmentation issue. This chapter contains two main sections; first, distinguishing overlapping cells from single cells, and second splitting the overlapping cells into single cells.

5.1 Methodology

To address the issue of overlapping cells, and consequently increasing the accuracy of cellular quantitative analysis a two stage algorithm is proposed: 1) distinguishing overlapping cells from single cells, and 2) splitting the identified overlapping cells into single cells. Figure 5.1 shows an overview of the developed algorithms.

Overlapping cells are distinguished from single cells by exploiting their different morphological properties. To split the identified overlapping cells into single cells, a novel algorithm is proposed which identifies their concave regions, initial splitting points, candidate pairs of initial splitting points and splitting route.

Five different tissue arrays were used to test and train the proposed system,

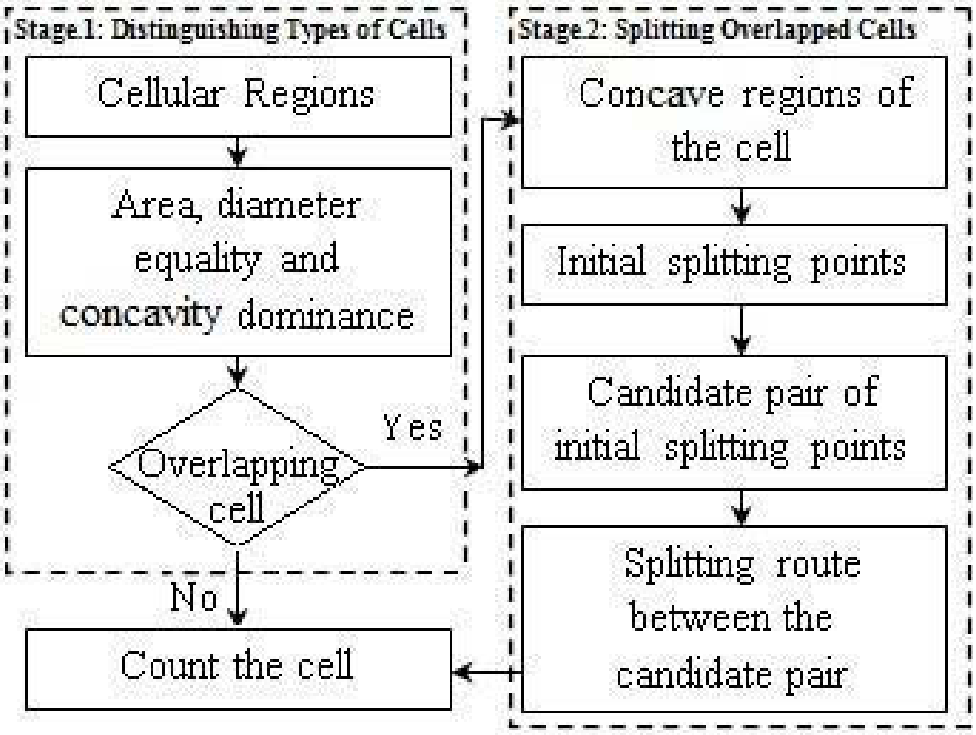


Figure 5.1: Overview of the two stages algorithm.

one tissue for training and the other four tissue arrays for testing the system. Each tissue array contained 50 tissue spots from different tumors with different neuroblastoma stages. The tissue arrays that are used for testing and training the algorithms in this chapter are different from those of the previous chapters. As described in Section 3.3 each of the tissue spots was cropped to an image of size 512×512 pixels with a JPEG format and a compression rate of 13:1. Thus, the test set contained four datasets each containing 50 images of neuroblastoma samples from different tumors.

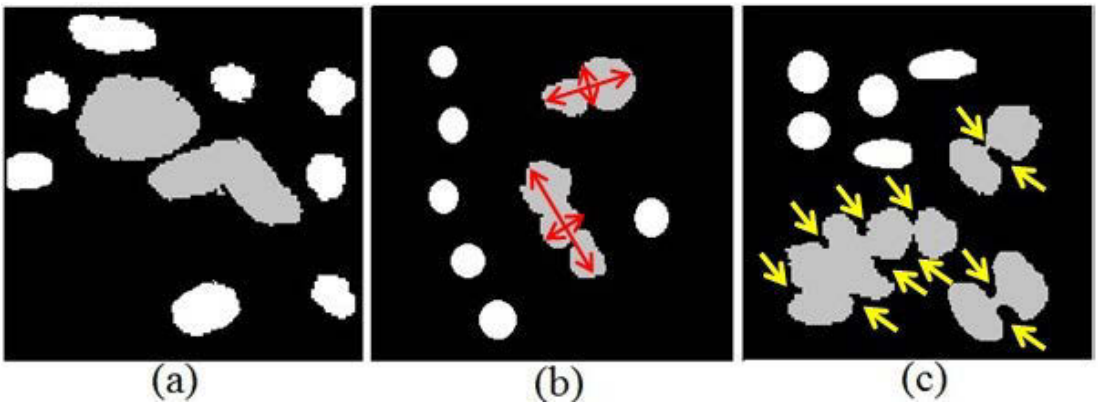


Figure 5.2: The gray cells are overlapped cells, and the white cells are single. a) Bigger area for overlapped cells. b) Overlapped cells with unequal diameters. c) Higher concavity for overlapped cells.

5.2 Proposed Approach

This section explains the algorithms which are proposed for distinguishing overlapping cells and splitting them into single cells in two subsections. The first subsection describes the developed algorithms for distinguishing overlapping cells from single cells. The second subsection describes the proposed algorithms for splitting overlapping cells.

5.2.1 Stage 1: Distinguishing between Types of Cells

To distinguish overlapping cells from single cells, pathologists consider certain morphological criteria of the cells such as size, diameter, and degree of cell concavity as shown in Figure 5.2. The method proposed in this thesis uses the same criteria to distinguish the types of cells as follows:

Area of the Cells

The size of overlapped cells is usually greater than the size of a single cell because an overlapped cell is formed by two or more cells. To distinguish overlapping cells

from single cells the proposed approach considers a cell as an object because it is not clear yet whether it is overlapping cell or single cell. Then, the area of each of the objects must be evaluated, and compared to the average size of the objects in the image. The area for each object is determined by counting the total number of constituent pixels within the cell, which is denoted as A_o, $\forall o \in \{1, ..., O\}$ where o is the object number, and O is the total number of objects in the image. Then the average size of the objects in each image is computed by $\overline{A} = \frac{1}{O}\Sigma_o^O A_o$.

Diameter Equality

A single neuroblast cell is roughly circular in shape. This means that the major and minor diameters of a single cell are approximately equal in length. The major diameter is the longest straight line inside the cell and is the cell's length. The minor diameter is a line perpendicular to the major diameter and is the cell width. The ratio of the minor diameter to the major diameter of object o is denoted by DE_o. If DE_o is bigger than a threshold γ, then object o is circular in shape. $\gamma = 0.9$ is obtained empirically by tuning on 50 images in the training set. The results of training indicate that different values for γ reduce the performance of the algorithm in detecting overlapping cells and that $\gamma = 0.9$ is the optimum value.

Concavity Dominance

Morphologically, overlapping cells have greater concavity than a single cell. The following expression is used to calculate the concavity of the objects:

$$CX_o = \frac{\hat{A}_o}{A_o} \tag{5.1}$$

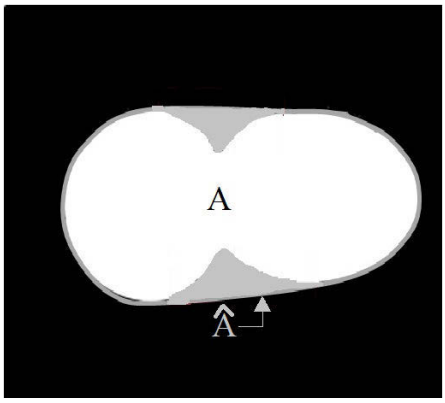


Figure 5.3: $\hat{A}$ and A of an overlapping cells.

where $\hat{A}_o$ is the area of convex hull for cell o, and CX_o indicates the cell concavity. Let us consider $\overline{CX}$ as the average cell concavity in the image. Then, $CX_o > \overline{CX}$ means that the cell has high concavity, otherwise the concavity is not high. Figure 5.3 indicates $\hat{A}$ and A and the concave regions with gray.

Labeling the Cells

To determine the type of object, a label is assigned to each of them. To do this, first the area of cells is checked using the following expression:

$$\zeta_o = \begin{cases} \text{PO}, & \text{if } A_o > \overline{A}, \quad \forall o \in \{1, ..., O\} \\ \text{PS}, & \text{Otherwise} \end{cases} \tag{5.2}$$

where PO and PS are Potentially Overlapping and Potentially Single labels respectively. In next step the ratio of minor to major diameters for each object is

computed, and new labels are assigned to the objects with respect to ζ_o using,

$$\zeta'_o = \begin{cases} \text{SC}, & \text{if } DE_o > \gamma \\ \text{PO}, & \text{if } DE_o \leq \gamma \text{ and } \zeta_o = \text{PO} \\ \text{PS}, & \text{if } DE_o \leq \gamma \text{ and } \zeta_o = \text{PS} \end{cases} \tag{5.3}$$

where SC is Single Cell label. The final step is to analyze the concavity of the objects with respect to ζ'_o and change the labels as following:

$$\zeta''_o = \begin{cases} \text{OC}, & \text{if } CX_o > \overline{CX} \text{ and } \zeta'_o = \text{PO or } \zeta'_o = \text{PS} \\ \text{SC}, & \text{Otherwise} \end{cases} \tag{5.4}$$

where OC is Overlapping Cell label. Ultimately, those objects with label $\zeta''_o = \text{OC}$ are considered as overlapping cells and the system stores them in a binary foreground/background image, where foreground is the overlapping cells with pixel value of 1, and the background is 0.

5.2.2 Stage 2: Splitting Overlapped Cells

The main markers that enable pathologists to identify the number of single cells in an overlapped cell are those areas which lie deeply inside the body of the overlapping cell. These areas form concave shapes in the body of the cells. To decompose an overlapped cell into its constituent single cells, pathologists draw an imaginary line from one concave region to the counterpart concave region to split the overlapped cells.

According to Figure 5.1 to automate the process of splitting overlapping cells, the proposed system first determines the concave regions of the cells. It then identifies initial splitting points on the concave regions, and selects the closest

initial splitting points as the candidate pair of initial splitting points. Finally, it determines the splitting routes using the shortest path between the selected candidate pairs.

Determining Concave Regions

Prior to applying the algorithm, a morphological opening operation (Soille, 2003) is applied on all of the overlapping cells. This operation removes all the small convex and concave regions which are not significant markers for identifying overlapping regions from the cell, and consequently, improves the performance of the algorithm in finding the correct concave regions.

To find the concave regions, R and $\hat{R}$ are assumed to be the entire region and the convex hull region of overlapping cells respectively. Then the regions between $\hat{R}$ and R are obtained using,

$$\phi_l = \hat{R}_l - R_l, \forall l \in \{1, ..., F\} \tag{5.5}$$

where F is the total number of overlapping cells. These regions have an approximately similar morphology to triangles and are denoted *Splitting Triangles* (STs) in this thesis as shown in Figure 5.4 by the hatched regions. Thus, $\phi_l = \{ST_x\}$, where $x = \{1, ..., H\}$ and H is the total number of STs for overlapping cell l. For example in this figure $H = 3$ because the overlapping cell contains three STs.

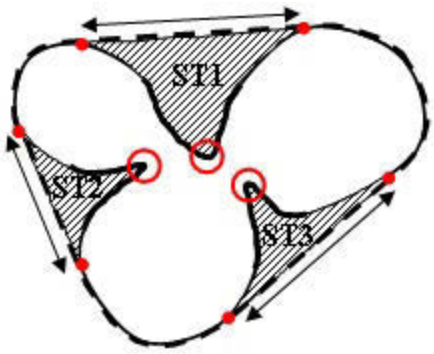


Figure 5.4: The dashed line indicates convex hull, the hatched regions are STs, the rings show the critical points on the arcs of the concave regions, the circles present critical points on the chords of STs, and the arrows point to the chords of STs.

Initial Splitting Points

To split overlapping cells, first the peak points of the regions which lie deeply inside the body of the cells must be defined as illustrated by the rings in Figure 5.4. These points are called the initial splitting points, because the algorithm initiates the splitting process from those points. These peak points are the critical points on the arcs of STs. Thus, the first step is to determine critical points of STs.

For each ST, three critical points are identifiable: Two points are on the chord of ST which are shown as circles and arrows in Figure 5.4, and one point is on the arc with a maximum distance from the chord. To split the cells, the critical point on the arcs are considered only, and the other two points are excluded from the analysis. To this end, a critical point detection algorithm based on concave region decomposition and seed growing techniques is introduced.

Initial Splitting Points: ST Decomposition

Each of the STs are decomposed into its edge, arc and chord. To determine the edge of the ST, Canny edge detection (Canny, 1986) is employed, and the

edge of an ST is denoted by $e(ST)$ as shown by dashed line in Figure 5.5b. The next step is to find the chord of the ST. The chord is the only part of ST which at a tangent to the contour of the cell convex hull as shown by the dashed line in Figure 5.5b. To determine the chord, the algorithm therefore uses the intersection of the edge of convex hull $e(\hat{R})$ of the cell and the edge of the ST using $chord_l(ST_x) = e_l(\hat{R}) \cap e_l(ST_x)$. Finally, to identify the arc of ST, the edge of ST is subtracted from the chord by $arc_l(ST_x) = e_l(ST_x) - chord_l(ST_x)$. The pixels on arcs and chords of all STs are stored in two matrices **EA** and **EC** respectively. The size of each of the matrices is $h \times w$ where h and w are equal to the height and the width of the original image respectively which is 512×512 in the experiments. In each of the matrices the pixels in the arc and the chord of cell l are set to l, and all other elements are set to 0.

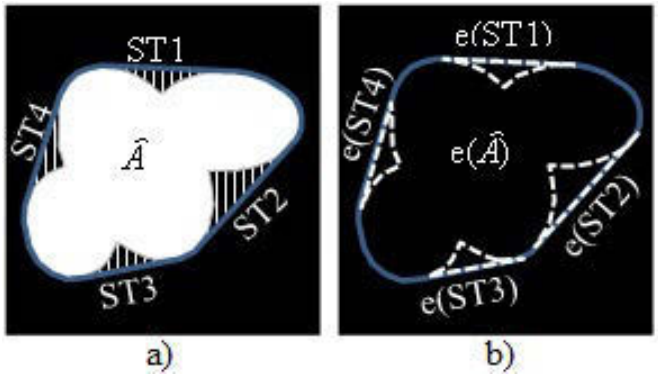


Figure 5.5: a) An example of the convex hull and STs of overlapping cells, b) illustrates the edge of the convex hull with the solid line; chords of STs with the dashed line which are tangent to the solid line, arcs with the dashed lines which are lie deeply in the body of the cell.

Initial Splitting Points: Seed Growing

The initial splitting point of an ST is located at the arc with the maximum distance from the chord, as shown in Figure 5.6a by the circle. Therefore, to identify the constituent pixels of an initial splitting point, the distance between pixels on

the arc and the chord must be computed. To this end, two distance matrices **DA** and **DC** with size $h \times w$ based on matrices **EA** and **EC** for arcs and chords are created respectively. Those elements of the matrices which have the same row and column numbers as the non-zero elements in **EA** and **EC** are considered 0, and are named 'seeds' in this thesis, and all other elements 'neighbors'. The values of the neighbors are assigned by growing from the seeds to the neighbors. That is, the distance between each neighbor and its nearest seed is calculated by the chessboard distance, $\max(|x_{neighbor} - x_{seed}|, |y_{neighbor} - y_{seed}|)$ (Cantrell, 2000) where $(x_{neighbor}, y_{neighbor})$ and (x_{seed}, y_{seed}) are the coordinates of a neighbor element and seed element respectively. Figures 5.6b and Figure 5.6c indicate matrices **DA** and **DC** with zero elements as the seeds and non-zero elements as the neighbors.

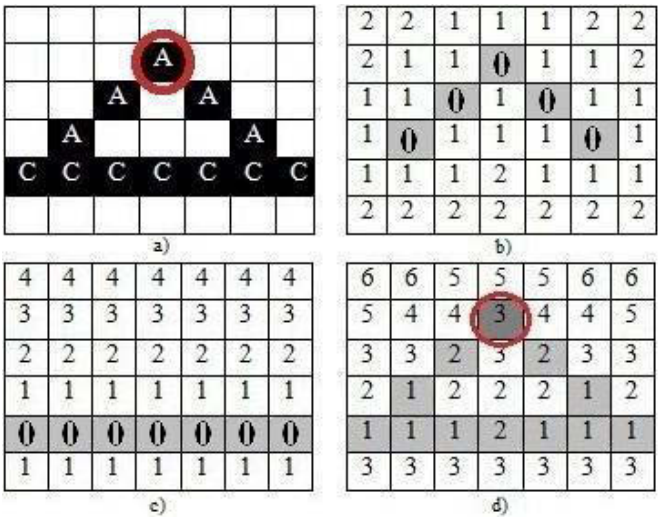


Figure 5.6: a) A and C represent the constituent pixels of an arc and a chord of an example ST, and the ring indicates the initial splitting point, b) indicates **DA** where all zero elements are seeds and all other elements are neighbors, c) illustrates **DC** with zero elements as seeds and neighbor elements, d) **DM**.

Initial Splitting Points: Critical Point Acquiring

The critical point of an arc is located at the maximum distance from the chord. To determine the critical point the following approach is developed:

$$\mathbf{DM} = \mathbf{DA} + \mathbf{DC} \tag{5.6}$$

Those elements in **DM** which align with the coordinates of non-zero elements in **EA** and have maximum values in **DM** are considered as the critical point of ST_x, and the initial splitting points. It means that each ST_x in overlapping cell l has an initial splitting point. Figure 5.6d shows an example **DM**.

Candidate Pairs of Initial Splitting Points

To split an overlapping cell, its nearest initial splitting points are connected to each other. Thus, the distance between the initial splitting points must be calculated, and two initial splitting points are taken as a pair if they are located nearest to each other. Two possible scenarios are considered:

Scenario 1: Overlapping cells with more than one ST: Figure 5.7 indicates an overlapping cell with more than one ST. In this scenario, each of the obtained initial splitting points is considered as an object, although an initial splitting point in an image might consist of more than one pixel. Then for each object, a border distance matrix $\mathbf{BD}_p$ is created with size $h \times w$, where $p = \{1, ..., SP\}$ and SP is the total number of objects (initial splitting points) for the overlapping cell l. To set the elements of the matrix, the developed seed growing technique is used by considering the constituent pixels of the object as the seed. Thus, the elements which have similar row and column numbers with the x and y coordinates of constituent pixels of the object in the image are set to 0, and the value of the

neighbors are set accordingly. Figure 5.8 is an example which shows the matrix $\mathbf{BD}_p$ for each of the identified initial splitting points.

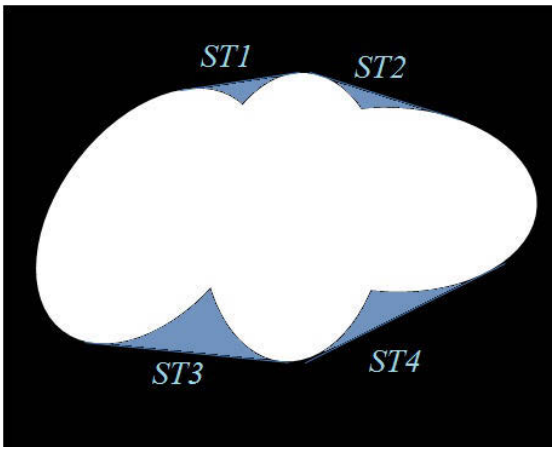
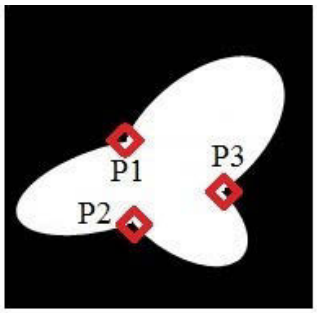


Figure 5.7: An example of overlapping cells with four *ST*s which are shown by gray regions. The red arrows indicate the distance between the initial splitting point to different points on contour of the overlapping cell.



2	1	1	1	1	2	3	4
2	1	**0**	**0**	1	2	3	4
2	1	1	1	1	2	3	4
2	2	2	2	2	2	3	4
3	3	3	3	3	3	3	4
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6

BD1

4	4	4	4	4	4	4	5
3	3	3	3	3	3	4	5
2	2	2	2	2	3	4	5
1	1	1	1	2	3	4	5
1	**0**	**0**	1	2	3	4	5
1	1	**0**	1	2	3	4	5
2	1	1	1	2	3	4	5
2	2	2	2	2	3	4	5

BD2

6	5	5	5	5	5	5	5
6	5	4	4	4	4	4	4
6	5	4	3	3	3	3	3
6	5	4	3	2	2	2	2
6	5	4	3	2	1	1	1
6	5	4	3	2	1	**0**	1
6	5	4	3	2	1	1	1
6	5	4	3	2	2	2	2

BD3

Figure 5.8: An example overlapping cell with three initial splitting points as shown by *P*s, matrix distances $\mathbf{BD_1}$ for P_1, $\mathbf{BD_2}$ for P_2, and $\mathbf{BD_3}$ for P_3, and **0**s are the constituent pixels of P_1, P_2 and P_3 in the cell.

To determine the shortest paths between two objects, the following four steps are developed:

1. For a selected object p, a set of matrices is introduced

$$\mathbf{\Theta}_p = \{\mathbf{B}_{p,v} = \mathbf{BD}_p + \mathbf{BD}_v \mid p, v \in \{1, ..., SP\}, p \neq v\} \tag{5.7}$$

where $\mathbf{\Theta}_p$ is a set of $SP-1$ matrices which indicates the distance between object p from the $SP-1$ other objects in an overlapping cell l as shown in Figure 5.9a.

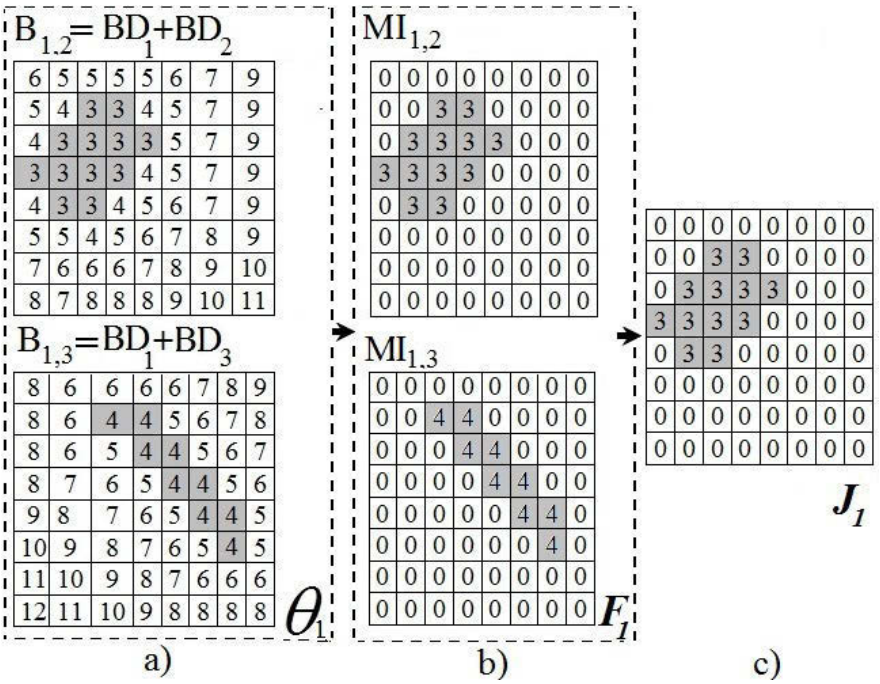


Figure 5.9: a) $\mathbf{BD}_1, \mathbf{BD}_2$ and $\mathbf{BD}_3$ are from Figure 5.8. b) $\mathbf{F}_1$ shows the shortest path between the objects, and c) $\mathbf{J}_1$ indicates that P1 and P2 in Figure 5.8 are the closest objects.

2. To obtain the shortest paths between object p and object v, a matrix $\mathbf{MI}_{p,v}$ is created with size $h \times w$ where all of its elements are 0 except those el-

ements which have the same row and column numbers to the elements in $\mathbf{B}_{p,v}$ with the minimum value.

3. $\mathbf{F_p}$ is a set of $SP-1$ matrices that have the smallest elements of the matrices in $\mathbf{\Theta}_p$ as illustrated in Figure 5.9b,

$$\mathbf{F}_p = \{\mathbf{MI}_{p,v} | v \in \{1, ..., s\}, p \neq v\} \tag{5.8}$$

4. The nearest object v to object p is determined by finding $\mathbf{MI}_{p,v}$ in set $\mathbf{F}_p$ which has the minimum value of greater than zero among the other matrices. The results are stored in $\mathbf{J}_p$ as shown in Figure 5.9c.

Scenario 2: Overlapping cells with one ST: Figure 5.10 indicates an overlapping cell with one ST. Sometimes an overlapping cell might have one concave region only which provides us with one initial splitting point. In this case, the system considers that the other initial splitting point to make the splitting pair is among those pixels which are located at the contour of the cell closest to the first initial splitting point. For this a two stage algorithm is developed:

1. Determine the edge of the overlapping cells. Canny edge detection is used to acquire the edge of the overlapping cells, and it is denoted as $e(R)_l$ where $l = \{1, ..., F\}$ and F is the total number of overlapping cells. Then $e(R)'_l = e(R)_l - arc(ST_1)_l$ is used to obtain all the edges of the overlapping cells except the edges which are in common to the concave region of the overlapping cell as shown by solid line in Figure 5.10.

2. Identify the shortest path between the initial splitting point of ST_1 and the constituent pixels of $e(R)'_l$ using the four steps from Scenario 1.

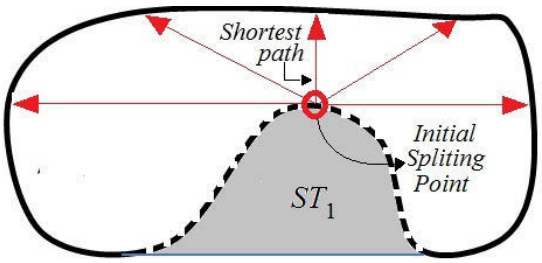


Figure 5.10: An example of overlapping cells with one ST only as shown by gray region.

Splitting Route Determination

The final part of stage 2 of the post-segmentation, as shown in the overview by Figure 5.1, is to find the splitting route between the candidate pairs. To this end, a single shortest path in $\mathbf{J}_p$ must be chosen. Consequently, the Thinning (Lam et al., 1992) morphological operation is applied to $\mathbf{J}_p$. Thinning is an iterative operation which reduces the width of the objects or the shape in the image. Two inputs can be considered for the Thinning operation, the points set, and the structuring element. The points set is the segmented image in this experiments, and the structuring element consists of a pattern specified as the coordinates of a number of discrete points relative to some origin. Figure 5.11 indicates the structuring elements that are used for the process of Thinning in this thesis.

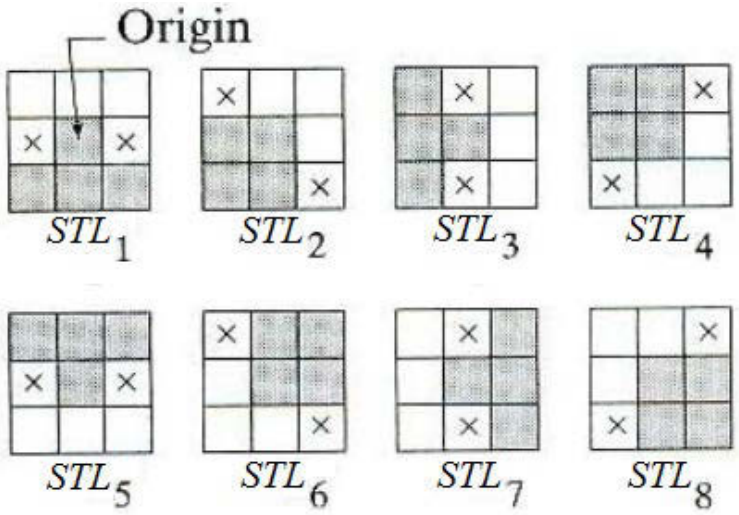


Figure 5.11: Eight different structuring element which are used for the Thinning process in this thesis.

Then the points set PS is thinned by structuring element STL using the expression $A \otimes \{STL\} = ((((A \otimes STL_1) \otimes STL_2) \ ...) \otimes STL_8)$ where $STL \in \{STL_1, STL_2, STL_3, STL_4, STL_5, STL_6, STL_7, STL_8\}$. This means that if the foreground and background of the points set fit the foreground and background of the structuring elements, the algorithm changes the value of the pixel underneath the origin of the structuring element to zero, otherwise it is left unchanged. This operation is repeated until no more changes occur. Usually the output is an object with the width of a single pixel. The results are stored in a matrix $\mathbf{J}_p^*$ with size $h \times w$. Figure 5.12 indicates $\mathbf{J}_p^*$ after applying the thinning operation. Finally in the segmented nuclei image, all the pixels with x and y coordinates equal to those non-zero elements are replaced in matrix $\mathbf{J}_p^*$ to 0. As a result, overlapping cells are split from their concave regions. Figure 5.14 shows an example of the proposed algorithm outputs.

The mathematical general view of the developed approach for splitting an overlapping cell is shown in figure 5.13.

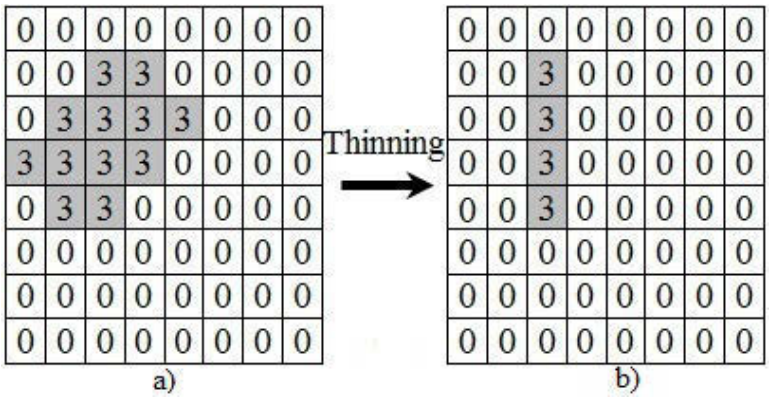


Figure 5.12: a) is matrix $\mathbf{J_1}$ in Figure 5.9c, and b) is $\mathbf{J}_1^*$.

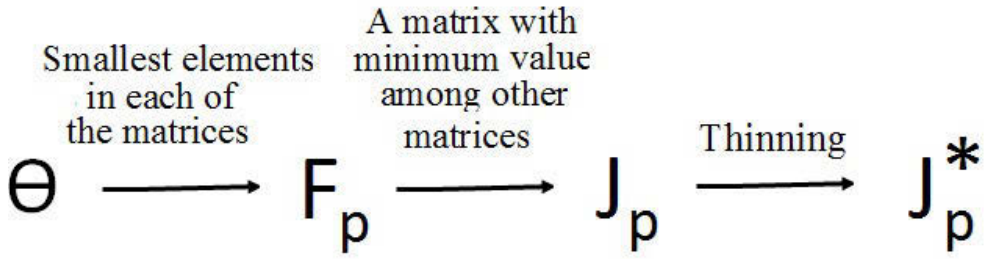


Figure 5.13: The mathematical view of the developed four steps for splitting an overlapping cells.

5.3 System Validation

Precision, recall and F-measure are used in this section to measure the difference between the results acquired by the proposed algorithm and those of the pathologist. A correctly segmented cell is True Positive (TP). A cell that has been detected by the system but rejected by the pathologist is False Positive (FP) or over-segmented, and a cell that has not been detected by the system but has been detected by the pathologist is False Negative (FN) or under-segmented. Performance of the system is compared to methods proposed by Kong et al. (2011), Zhou and Li (2009), and the Watershed technique by Roerdink and Meijster

(2000) as reviewed in Section 2.2.1. Four test datasets as described in Section 5.1 are used for this comparison.

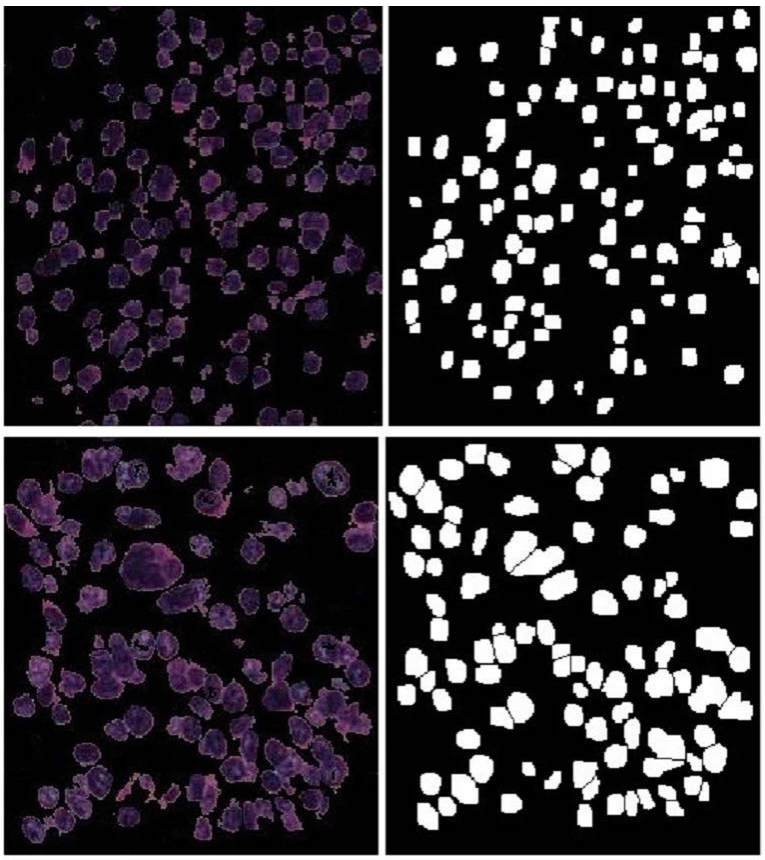


Figure 5.14: The left two images are segmented nuclei images and the right binarised images show the outputs of the algorithm after applying to the segmented nuclei images.

5.4 Experimental Results

Each column of Table 5.1 indicates the average TP, FP, FN, precision, recall and F-measure of the developed algorithms for each test dataset after comparison with the pathologist.

Table 5.1: The average TP, FP, FN, precision, recall and F-measure obtained by the proposed post-segmentation algorithms in this thesis.

Data	TP	FP	FN	Precision	Recall	F-measure
Dataset.1	106.2	14.95	11.55	87.66%	91.00%	89.29%
Dataset.2	146.83	23.57	17.19	86.16%	89.52%	87.80%
Dataset.3	181.98	25.11	23.31	87.87%	88.64%	88.25%
Dataset.4	105.64	12.02	12.91	89.78%	89.11%	89.44%
Average				87.86%	89.56%	88.70%
StdDev				1.48	1.02	0.79

Table 5.2: The average F-measure obtained by the proposed system in this thesis and state of the art systems.

Data	Our System	Kong et al.	Zhou et al.	Watershed
Dataset.1	89.29%	83.38%	82.75%	79.40%
Dataset.2	87.80%	85.37%	84.68%	75.29%
Dataset.3	88.25%	82.32%	82.28%	77.23%
Dataset.4	89.44%	86.13%	83.72%	74.71%
Overall	88.70%	84.30%	83.36%	76.66%
StdDev	0.79	1.75	1.07	1.83

Table 5.2 indicates the results obtained by the proposed system in this thesis and the proposed systems by Kong et al. (2011), Zhou and Li (2009), and the Watershed technique (Roerdink and Meijster, 2000) after applying to the four test datasets. According to the table the proposed systems has the best average F-measure followed by Kong et al. with an average F-measure of 84.30%. This is the best result of the state of the art methods, but is approximately 4% lower

than the F-measure which obtained by the system in this thesis. The lowest F-measure of 76.66% is obtained by Watershed.

According to Table 5.2, after the Watershed technique, the system proposed by Kong et al. has the highest standard deviation, which indicates that these systems have fluctuating performance for different images, and are sensitive to noise and image quality. Proposed system by Kong et al. analyzes the cell's contours to detect and split the overlapping cells, and the main reason for the high standard deviation is the sensitivity of the system to the fluctuation at the contour of cells. The proposed system in this thesis has the lowest standard deviation, which indicates low sensitivity to noise.

Tables 5.3 and 5.4 indicate that the system proposed in this thesis obtains the best precision and recall, while the watershed technique achieves the worst result. The low precision for watershed is indicative of over-segmentation. As explained in Section 2.2, Zhou and Li (2009) addresses the over-segmentation drawback of the watershed technique by merging the fragmented cells which is why the system proposed by Zhou and Li (2009) is better than watershed. Although Zhou and Li (2009) system significantly improves the performance of the watershed technique, the lowest precision results of the watershed technique is indicative of higher over-segmentation in comparison with system proposed by Kong et al. and the system in this thesis.

Table 5.3: The average precision obtained by the proposed system in this thesis and state of the art systems.

Data	Our System	Kong et al.	Zhou et al.	Watershed
Dataset.1	87.66%	82.67%	81.46%	77.11%
Dataset.2	86.16%	84.08%	84.61%	72.84%
Dataset.3	87.87%	82.19%	80.12%	75.32%
Dataset.4	89.78%	84.51%	82.41%	73.02%
Overall	87.86%	83.36%	82.15%	74.57%
StdDev	1.48	1.10	1.88	2.03

Table 5.4: The average recall obtained by the proposed system in this thesis and state of the art systems.

Data	Our System	Kong et al.	Zhou et al.	Watershed
Dataset.1	91.00%	84.11%	84.08%	81.83%
Dataset.2	89.52%	86.71%	84.77%	77.92%
Dataset.3	88.64%	82.46%	84.57%	79.25%
Dataset.4	89.11%	87.83%	85.08%	76.49%
Overall	89.56%	85.27%	84.62%	78.87%
StdDev	1.02	2.44	0.42	2.27

Table 5.5 indicates the results of Holms statistical tests which were applied to demonstrate that the algorithm proposed in this thesis significantly outperforms the above state of the art algorithms in splitting overlapping cells. The null hypothesis of the tests is given as similarity between precision, recall and F-measure of the algorithm proposed in this thesis and those of the algorithm proposed by Kong et al. (2011). The algorithm proposed by Kong et al. (2011) is chosen for the statistical tests among the other algorithms, because it obtained the best performance among other state of the art algorithms in Tables 5.2, 5.3

and 5.4. Table 5.5 shows that p-values for precision, recall and F-measure for each of the four test datasets are far less than the level of significance, and consequently the hypothesis is rejected. These results conclude that the algorithm proposed in this thesis is significantly better than the algorithm proposed by Kong et al. (2011), and the other algorithms in the experiments.

Table 5.5: Holms statistical tests for precision, recall and F-measure of the algorithm proposed in this thesis and the algorithm proposed by Kong et al. in splitting overlapping cells with $\alpha = 0.1$.

Dataset	Metric	z_0	p-value
Dataset 1	Precision	4.024	5.699×10^{-5}
	Recall	4.024	5.699×10^{-5}
	F-measure	3.13	0.001
Dataset 2	Precision	2.683	0.007
	Recall	2.459	0.01
	F-measure	3.13	0.001
Dataset 3	Precision	4.248	2.151×10^{-5}
	Recall	3.801	1.439×10^{-4}
	F-measure	3.801	1.439×10^{-4}
Dataset 4	Precision	4.248	2.151×10^{-5}
	Recall	3.130	0.001
	F-measure	2.459	0.01

5.5 Related Publications

All of the proposed algorithms in this chapter are published in peer-reviewed journals and conference papers.

Refereed Journal Papers,

1. S. Tafavogh, K. Felix Navarro, D. R. Catchpoole, and P. J. Kennedy. "Non-parametric and integrated framework for segmenting and counting neuroblastic cells within neuroblastoma tumor images". *Medical and Biological Engineering and Computing*, Springer, 51:645–655, 2013. (PMID: 23359256).

2. November 2013: submitted to journal of IEEE Transactions on Information Technology in Biomedicine, S. Tafavogh, D. R. Catchpoole and P. J. Kennedy, "Cellular quantitative analysis and splitting overlapping cells using shortest path between critical points on the concave regions".

Refereed Conference Paper,

1. S. Tafavogh, K. Felix Navarro, D. R. Catchpoole and P. J. Kennedy (2013). "Segmenting cellular regions of neuroblastoma tumor and splitting overlapping cells using shortest path between convex regions of cell contours", In Artificial Intelligence in Medicine, Springer, LNCS 7885, N. Peek, R. M. Morales and M. Peleg, Eds, Springer. 171-175.

5.6 Contribution, Discussion and Conclusion

The proposed algorithms in this chapter address Contribution 3 of this thesis which is to identify and split overlapping cells. Overlapping cells in the sample of tumor tissues is the major issue in decreasing the accuracy of automated cellular quantitative analysis systems. This chapter addresses the issue of overlapping cells for the cellular quantitative analysis by proposing a novel post-segmentation algorithm. The algorithm contains two stages: 1) distinguishing overlapping cells from single cells based on their morphological differences, and 2) splitting the

identified overlapping cells into single cells using the proposed splitting triangle decomposition, seed growing and shortest path determination techniques. The results obtained by the proposed algorithms indicate superior performance compared to robust state of the art methods. The results obtained indicate that the proposed algorithms provide high accuracy in identifying overlapping cells and splitting them into single cells. The standard deviation obtained by the proposed system, which was the lowest compared to other systems demonstrating the low sensitivity of the system to noise.

Chapter 6

Neuroblastoma Classification: Quantitative Analysis

This chapter addresses **Contribution 4** of this thesis by proposing a prognosis decision making engine to grade the level of aggressiveness for NTs. To measure the tumor aggressiveness, the level of tumor maturation must be determined. The quantity and the quality of different types of histological structures are the main markers for classifying maturation of NTs. In Section 2.1 different levels of tumor aggressiveness were described in detail. Shimada et al. (1999) introduces four types for NT to determine the aggressiveness of the tumor: 1) neuroblastoma, 2) ganglioneuroblastoma intermixes, 3) ganglioneuroma maturing and mature and 4) ganglioneuroblastoma nodular. In this thesis the neuroblastoma type is taken into consideration because the other types require macroscopic analysis in addition to microscopic analysis and this is beyond the scope of this thesis. Neuroblastoma has three subtypes including undifferentiated, poorly differentiated and differentiated as shown in Figure 6.1.

To determine the subtypes of neuroblastoma, the Shimada classification intro-

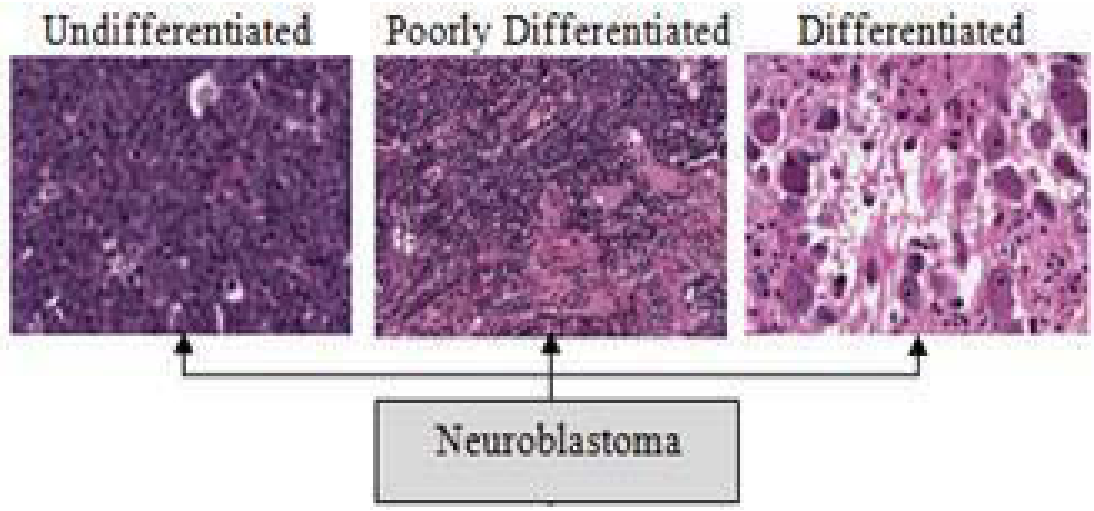


Figure 6.1: Shimada classification of neuroblastic tumor

duces a set of rules based on the amount of neuropil, ganglion cells and neuroblast cells within the tissue of the tumor. This thesis proposes a set of computerized rules according to the Shimada classification to classify neuroblastoma into undifferentiated, poorly differentiated and differentiated subtypes.

6.1 Methodology

In order to perform the qualitative analysis on the NT images and to grade the tumor, the system must first localize and segment the key histological structures, and then apply computerized rules to determine the grade of the tumor. Chapters 4 and 5 developed and validated approaches for localizing and segmenting each of the histological regions and histological structures including neuropil region, ganglion cells and neuroblast cells. This chapter proposes a prognosis decision making engine using a set of rules based on the Shimada classification to use the segmented histological structures for grading NTs. Figure 6.2 indicates a schematic view of the qualitative analysis in this chapter.

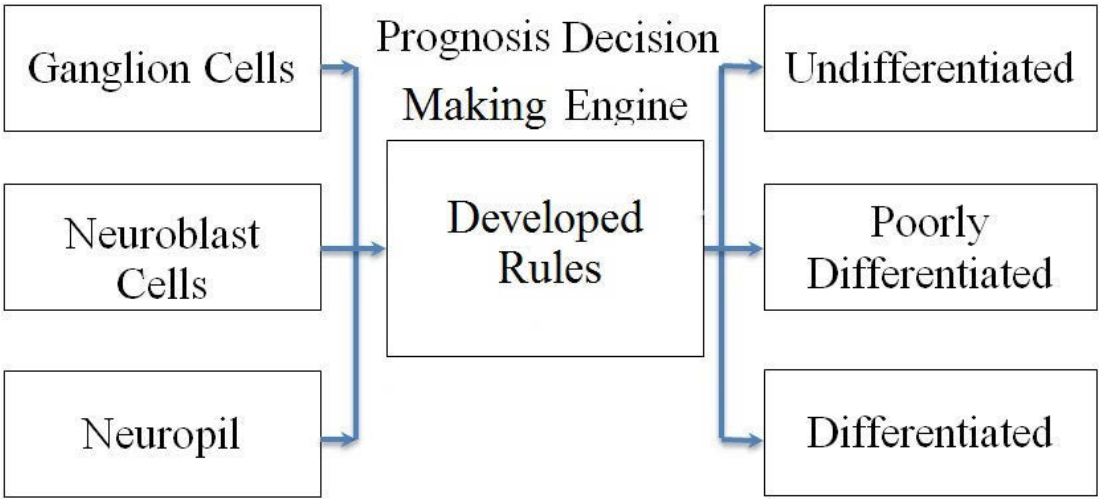


Figure 6.2: Schematic view of the qualitative analysis.

A test set containing 81 images of the tissue spots are used to validate the performance of the prognosis decision making engine. The test set is made up of 27 of each of the subtypes. The NT grade for each of the images was specified by the pathologist.

6.2 Experimental Approach

6.2.1 Prognosis Decision Making Engine

The prognosis decision making engine analyzes the information obtained from the segmented histological regions and structures, and uses a set of computerized rules to grade the aggressiveness level of NTs. To this end, the system must determine the amount of each of the histological structures in the image of the tumor.

To determine the amount of neuropil regions within the image, the engine calculates the ratio of neuropil regions to the addition of neuropil regions to

cellular regions as following expression:

$$NR = \frac{np}{cp + np} \tag{6.1}$$

where np is the total number of constituent pixels within the neuropil regions and cp is the total number of constituent pixels within the cellular regions. The ratio of ganglion cells to all other cell types is determined by:

$$GR = \frac{ganglion}{ganglion + neuroblast} \tag{6.2}$$

where *ganglion* is the total number of identified ganglion cells in the segmented cellular regions and *neuroblast* is the total number of neuroblast cells within the cellular regions.

A set of rules is defined according to Shimada et al. (1999) to classify NTs using the obtained NR and GR. Figure 6.3 indicates the schematic view of the rules proposed for grading the NT.

According to the Shimada scheme if more than 2% of a tumor consists of neuropil regions, the tumor is either poorly differentiated or differentiated, otherwise it is undifferentiated. Thus, the CAD must analyze the amount of neuropil regions in the tumor. As a result, Rule 1 is proposed to analyze the value of NR, and if $NR \leq 2\%$, the tumor then is considered as undifferentiated, otherwise it deploys Rule 2. Shimada scheme considers the amount of ganglion cells in the tumor to distinguish between poorly differentiated and differentiated. According to Shimada scheme if more than 5% of tumor cellular regions consist of ganglion cells, the tumor is differentiated, otherwise it is poorly differentiated. Rule 2 of the system is developed based on the same rationale for distinguishing between

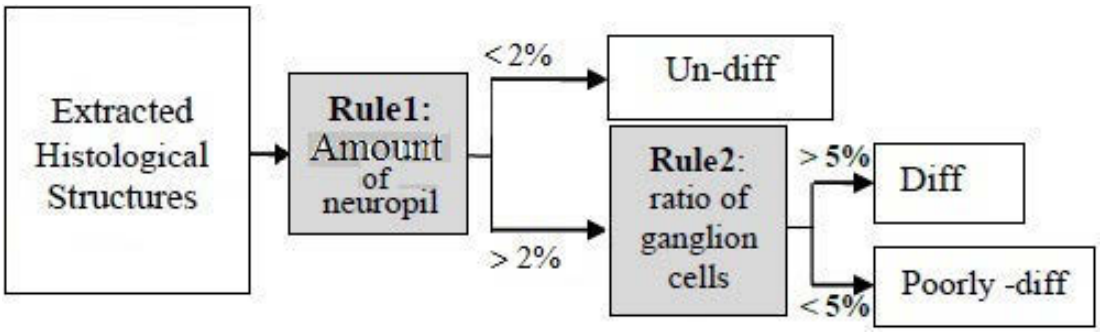


Figure 6.3: Schematic view of the defined rules for the system.

Algorithm 1 Classifying NT using microscopic criteria

1: **Rule 1:**
Apply expression (6.1) to determine the amount of NR.
If $NR < 0.02$, **then** NT$\longrightarrow$ Undifferentiated.
Otherwise, Rule2.
2: **Rule 2:**
Apply expression (6.2) to determine the amount of GR.
If $GR < 0.05$ **then** NT$\longrightarrow$ Poorly Differentiated.
Otherwise, NT$\longrightarrow$ Differentiated.

poorly differentiated and differentiated. It means that if $GR > 5\%$ the tumor then is differentiated otherwise it is poorly differentiated. Algorithm 1 presents all the steps which are proposed to classify NT based on the combination of the above-stated microscopic criteria.

6.3 System Validation

To evaluate the performance of the prognosis decision making engine, it is applied to the 81 test images. Also each of these test tissue images were manually analyzed and their subtypes were determined by the pathologist. The performance of the system in grading NTs is then measured by comparing the obtained results by the CAD with those of the pathologist as the ground truth. Moreover, to evaluate performance of the system, it is compared to the only other automated

qualitative analysis state of the art system which is the one proposed by Kong et al. (2009) for determining the grades of NTs. The system proposed by Kong et al. is redeveloped based on the algorithms introduced in Kong et al. (2009) for comparing with the system proposed in this thesis. The detail of the algorithms are explained in Section 2.2.2.

6.4 Experimental Results

Table 6.1 indicates the confusion matrices of the proposed system in this thesis and the system proposed by Kong et al. Table 6.2 indicates the average precision, recall and F-measure of both systems based on the results shown in Table 6.1. The results indicate that the proposed system obtains higher precision, recall and F-measure in comparison with that of Kong et al.

The most aggressive subtype of Neuroblastoma is undifferentiated, then poorly differentiated and lastly differentiated, and the proposed method by Kong et al. wrongly classifies 8 tumors from 27 tumors of undifferentiated subtypes to other subtypes of Neuroblastoma. This means that the system suggests to biologists to use a different treatment for an aggressive subtype of NT.

The main reason for big differences between the systems is that the system proposed in this thesis uses all the histological regions and histological structures which are used by pathologists to grade the NTs. As mentioned in Section 2.2.2, the system proposed by Kong et al. uses a combination of classifiers based on segmented nuclei, cytoplasm and background regions to make the decision for the grade of NT. However, these features are not key markers for grading the NTs by pathologists.

Table 6.1: Confusion matrices for the system developed in this thesis and the system proposed by Kong et al. (2009). UD = Undifferentiated, PD = Poorly Differentiated, and D = Differentiated.

Actual	Prediction		
	UD	PD	D
UD	25	1	1
PD	1	24	2
D	0	1	26

Proposed system in this thesis.

Actual	Prediction		
	UD	PD	D
UD	19	3	5
PD	8	15	4
D	3	2	22

Proposed system by Kong et al.

Table 6.2: Average precision, recall and F-measure for the system proposed in this thesis and the system proposed by Kong et al. (2009).

System	Precision	Recall	F-measure
The proposed System	92.59%	92.80%	92.69%
Kong et al.	69.13%	69.76%	69.44%

6.5 Contribution, Discussion and Conclusion

This chapter addresses Contribution 4 of this thesis by developing algorithms for the automated qualitative analysis of images of NTs. The algorithms proposed

in this chapter take the outputs of other chapters for determining the stage of NTs. It means that the outputs of all the techniques developed for extracting the histological regions and histological structures within NTs are integrated and processed by the set of rules for determining the stage of NTs. The proposed computerized rules in this thesis are developed based on the Shimada classification. Results obtained by the proposed system and the other state of the art system were compared with those of a pathologist. The results indicate that the performance of the system developed in this thesis is superior to the method proposed by Kong et al. The main reason for this superior result is that the proposed system takes the same histological regions, histological structures, and classification method in to consideration as pathologists do. The system proposed by Kong et al. (2009) does not take all the necessary histological structures and abnormalities into consideration. Moreover, the computerized rules proposed by Kong et al. (2009) which are used for grading the NTs are not based on a histopathology classification.

The next step after this successful proof of the concept is to compare the performance of the CAD in their grading NTs to the performance of pathology laboratories with inter-intra observer variability in grading. If the CAD obtains higher performance in grading NTs compared with the performance of the laboratories, it then can be used as a second reviewer for pathology laboratories. According to Arbiser et al. (2001), consultative review of pathology materials significantly enhances the quality assurance and accuracy of diagnosis for cancers and 65% of institutions (international) apply the consultation review on their cases only. Thus, the CAD has the potential to increase the accuracy of pathology laboratories and consequently improve patients chance of survival by serving as the second reviewer.

Chapter 7

Conclusions

Conventional pathology has some degree of inter-intra observer variability in cancer grade determination. Moreover, conventional pathology is not well aligned with advances in technology, and still the optical microscopes and glass slides are the main tools for determining diagnosis and prognosis for cancers. The discordance between the prototypes proposed by central pathologists and the diagnosis/prognosis of the institutional pathologists reduces the chance of survival and increases the medical error rate. Thus, there is an urgent demand from pathology departments for a framework and system which reduces the discordance and improves the chance of survival for the patients.

Digital pathology, as a new framework for conventional pathology, assists pathologists by the use of computerized analysis of tumor tissues to increase the efficiency of diagnosis, prognosis and treatment of the cancerous tumor. A Computer Aided Diagnosis system is a principle component of digital pathology and is becoming a part of routine clinical work.

A Computer Aided Diagnosis system deploys computerized algorithms to perform quantitative and qualitative analysis on the histological images of cancerous

tumor tissues. But still many of the CAD systems fail to provide an accurate and reliable quantitative or qualitative result to pathologists. Especially for neuroblastoma, a CAD system that integrates both quantitative and qualitative analysis accurately and follows the histopathology scheme for grading the cancer has not been proposed. The system proposed by Kong et al. (2009) is the only system which performs quantitative and qualitative analysis on the tumor of NTs, but the main drawback is that it does not follow a specific histopathology scheme.

The research in this thesis is motivated by the fact that most of the CAD systems suffer from a lack of efficiency for four major reasons as reviewed in the literature: 1) The fluctuating quality of the histological images. 2) A wide range of types of histological structures, abnormalities and histological regions each requiring its own adopted algorithm. 3) Low accuracy in quantitative analysis of the cellular regions, because of the overlapping cells. 4) A lack of utility because systems do not follow the required histopathological schemes for grading cancers. In light of these issues, this research makes the following main contributions:

Contribution 1 proposes a novel two-step algorithm in Chapter 4 to reduce the wide range of intensity variation between the constituent pixels of the histological images. The first step of the algorithm is a global intensity variation reduction technique. By using this technique the range of pixel intensities in the entire image is reduced. The second step reduces the intensity variation between constituent pixels of homogeneous regions in terms of intensity and spatial domain. These homogeneous regions form one type of histological region. To achieve this, a local intensity variation reduction technique is proposed. The rationale behind this technique is to reduce the intensity variation between the constituent pixels of a histological region using the proposed three step operation: 1) color space transformation, 2) mosaicking the image, and 3) intensity unification.

Contribution 2 introduces a series of segmentation algorithms in Chapter 4 for locating and extracting different histological regions, abnormalities and histological structures within the histological images of NTs. The proposed algorithms locate cellular regions and neuropil regions, segment nuclei and cytoplasm of the cellular regions, distinguish ganglion cells from neuroblast cells and extract them, and count the total number of ganglion cells and neuroblast cells. In order to improve the performance of the system, the pre-segmentation and segmentation techniques are integrated. The performance of the algorithms was evaluated by a pathologist. Moreover, the performance of the proposed system in this thesis was compared with those of state of the art systems, and the results demonstrate statistical superiority for the proposed algorithms in this thesis over the state of the art systems.

Contribution 3 develops a robust post-segmentation algorithm in Chapter 5 for discriminating overlapping cells from single cells and for splitting the overlapping cells into single cells. Errors are derived from the inability of most of the reviewed CAD systems to distinguish between overlapping cells and single cells, and to split the overlapping cells. An enormous number of overlapping cells in the cancerous tissues makes the process of cell counting error-prone. In this thesis the issue is addressed by: 1) Analyzing the morphological differences between overlapping cells and single cells using area, minor and major diameters, and concavity of the cells to distinguish between overlapping and single cells. 2) Identifying the initial splitting points by deploying the developed concave regions determination, decomposition, seed-growing and critical point acquiring techniques. Finally, a shortest path determination method is developed to identify a splitting route between the candidate pair of initial splitting points. Thus the proposed algorithm splits the segmented overlapping cells into single cells. This

process improves the performance of the system in cellular quantitative analysis of tissue. The performance of the algorithms was evaluated by a pathologist. Moreover, the performance of the proposed system in this thesis was compared with those of state of the art systems, and the results show statistical superiority for the proposed algorithms in this thesis over the state of the art systems.

Contribution 4 proposes an automated prognosis decision engine in Chapter 6 to perform the qualitative analysis of NT histological images. The engine is designed based on the Shimada histopathology scheme which is one of the most comprehensive and well established grading schemes for NTs. To grade the NTs, the system takes into account the same morphological features and histological structures as the Shimada scheme. The advantages of this approach is that in contrast to the only competing state of the art CAD system for NTs which make diagnosis and prognosis according to computerized classification techniques, the system proposed in this thesis makes a decision about the grade of NT based on a robust histopathology scheme. Thus, the performance of the system in diagnosis and prognosis is more assessable and readable for pathologists, and it consequently increases the utility of the system by meeting the clinical demands of a real-life situation.

7.1 Discussion

The proposed CAD system has a number of potential benefits for the health sector and for computer science. In the health sector, the proposed CAD system assists pathologists in prognosis and making appropriate treatment by targeting the most destructive infant cancer. It facilitates the process of decision-making for the prognosis of NT by addressing one of the most important prognostic factors

which is grading the aggressiveness of the tumor. In terms of computer science, the proposed methodology, framework and algorithms can be used as a platform for developing other types of CAD systems.

The proposed CAD system assists pathologists with one of the most important factors of NT prognosis which is determining whether the grade of the tumor is undifferentiated, poorly differentiated or differentiated. However the proposed CAD does not provide a report about the favorable or unfavorable histology of the tumor. In order to determine this a number of specific types of nerve cells such as mitotic and karyorrhectic cells must be located and segmented by the CAD. The other limitation for this research is the macroscopic analysis of NTs. Macroscopic analysis means that pathologists analyze the original tumor tissue without using the microscope to obtain information such as the size of the malignant tumor, hemorrhage symptoms, the texture of the tumor, and abnormality in the tumor. This data can not be transformed into a digital format until a clinician enters it into the CAD. This process does not reduce any of the work load for pathologists and it is prone to human errors.

7.2 Future Study

As explained in the discussion, determining favorable or unfavorable histology of the tumor assists pathologists in determining a patient's chance of survival. Patients with a favorable histology have a higher chance of survival than patients with an unfavorable histology. According to Shimada et al. (1999) determining a favorable or unfavorable histology depends on the ratio of two specific types of cells called mitotic cells and karyorrhectic cells. Mitotic cells are those in the process of regeneration by dividing their nuclei. They have similar morphologi-

cal features to overlapping cells. The karyorrhectic cells are those which contain fragmented nuclei. The ratio of mitosis and karyorrhexis to other cellular components in the tissue is known as the Mitotic Karyorrhectic Index (MKI). The three classes considered for MKI are high, intermediate and low. The combination of information such as the grade of NTs and the class of MKI indicates a patient's chance of survival. For example, an undifferentiated subtype of tumor with high MKI is an indicator of unfavorable histology, while a differentiated subtype of NT with a low MKI signifies a better patient outcome. Therefore, determination of the MKI would enhance the CAD proposed in this thesis to better determine the patient's chance of survival.

To determine MKI, the proposed quantitative and qualitative analysis approaches in Chapters 4 and 6 respectively could be extended to determine more histological regions, histological structures and computerized rules. Moreover, to improve the dissemination of the system Cloud technology can be used. As a result our future suggested directions in this research are summarized in the following tasks:

Extending quantitative analysis of the CAD to segment mitotic and karyorrhectic cells. Mitotic cells can be located and segmented by incorporating the proposed concave region analysis in this thesis with cell contour analysis and edge detection techniques. Due to the specific texture of the Karyorrhectic cells, they can be located and segmented using texture analysis.

Extending qualitative analysis of the system by improving the computerized rules to determine patient's chance of survival by integrating the MKI rate with the aggressiveness grade of the NT. Thus new rules could be added to the prognosis decision engine in Chapter 6 for counting the total number of obtained mitotic and karyorrhectic cells to determine the MKI class, and for integrating

the results with the aggressiveness grade of the NTs.

Enhancing dissemination of the system to provide an economical and easily accessible CAD for the health sector. The system could be available online using the Cloud technology. The CAD could be located on clusters of high performance computers and all the processes and computations performed by them. Pathology laboratories connect to the CAD using Internet facilities. The advantages of using Cloud technology and online systems are: 1) pathology labs will not be dependent on specific hardware components and all the hardware tools will reside in the server; 2) there will be no need to install or maintain software; 3) the system will be accessible to pathologists wherever Internet facilities are provided; 4) all the computational processes will be performed on the Cloud and pathologists will only need to upload an image of the regions of interest to the Internet through their tablets, or other computer devices. Of course potential legal and ethical issues must to be overcome. Moreover, in order to introduce the CAD system to pathologists, the running of a trial version of the CAD in pathology laboratories must be undertaken.

Finally, applying a number of modifications to the proposed algorithms allows us to use the CAD system for other cancers where counting cells is one of their prognostic factors.

The hypothesis in this thesis was to provide a robust and accurate CAD system which performs quantitative and qualitative analysis on the NTs. The results demonstrate that the proposed CAD system for automated quantitative and qualitative analysis is not only feasible, but it statistically significantly outperforms all other state of the art methods. The next steps are to extend the algorithms for determining the MKI rate, and ultimately to deploy the CAD in pathology laboratories and hospitals. This is an exciting time for CAD systems and their

use in hospitals for addressing the needs of pathologists.

Bibliography

Al-Kofahi, Y., Lassoued, W., Lee, W. and Roysam, B. (2010), 'Improved automatic detection and segmentation of cell nuclei in histopathology images', *IEEE Transactions on Biomedical Engineering* **57**(4), pp. 841–852.

Arbiser, Z. K., Folpe, A. L. and Weiss, S. W. (2001), 'Consultative (expert) second opinions in soft tissue pathology analysis of problem-prone diagnostic situations', *American Journal of Clinical Pathology* **116**(4), pp. 473–476.

Baade, P., Youlden, D., Valery, P., Hassall, T., Ward, L., Green, A. and Aitken, J. (2010), 'Trends in incidence of childhood cancer in Australia, 1983–2006', *British Journal of Cancer* **102**(3), pp. 620–626.

Bai, X., Sun, C. and Zhou, F. (2009), 'Splitting touching cells based on concave points and ellipse fitting', *Pattern Recognition* **42**(11), pp. 2434–2446.

Bamford, P. and Lovell, B. (1998), 'Unsupervised cell nucleus segmentation with active contours', *Signal Processing* **71**(2), pp. 203–213.

Beckwith, J. B. and Martin, R. F. (1968), 'Observations on the histopathology of neuroblastomas', *Journal of Pediatric Surgery* **3**(1, Part 2), pp. 106 – 110.

Begelman, G., Gur, E., Rivlin, E., Rudzsky, M. and Zalevsky, Z. (2004), Cell

nuclei segmentation using fuzzy logic engine, *in* 'International Conference on Image Processing, 2004. ICIP'04.', Vol. 5, IEEE, pp. pp. 2937–2940.

Beucher, S. and Lantuejoul, C. (1979), Use of Watersheds in Contour Detection, *in* 'International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.'.

Boucheron, L. E., Manjunath, B. and Harvey, N. R. (2010), Use of imperfectly segmented nuclei in the classification of histopathology images of breast cancer, *in* 'IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)', IEEE, pp. 666–669.

Buades, A. et al. (2011), 'Self-similarity-based image denoising', *Communications of the ACM* **54**(5), pp. 109–117.

Canny, J. (1986), 'A computational approach to edge detection', *IEEE Transactions on Pattern and Machine Intelligence* (6), pp. 679–698.

Cantrell, C. D. (2000), *Modern mathematical methods for physicists and engineers*, Cambridge University Press.

Coelho, L. P., Shariff, A. and Murphy, R. F. (2009), Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms, *in* 'IEEE International Symposium on Biomedical Imaging, ISBI'09.', IEEE, pp. 518–521.

Colantonio, S., Salvetti, O. and Gurevich, I. (2007), 'A two-step approach for automatic microscopic image segmentation using fuzzy clustering and neural discrimination', *Pattern Recognition and Image Analysis* **17**(3), pp. 428–437.

Comaniciu, D. and Meer, P. (2002), 'Mean shift: a robust approach toward feature space analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5), pp. 603–619.

Cong, G. and Parvin, B. (2000), 'Model-based segmentation of nuclei', *Pattern Recognition* **33**(8), pp. 1383–1393.

Cooper, L., Sertel, O., Kong, J., Lozanski, G., Huang, K. and Gurcan, M. (2009), 'Feature-based registration of histopathology images with different stains: an application for computerized follicular lymphoma prognosis', *Computer Methods and Programs in Biomedicine* **96**(3), pp. 182–192.

Cornish, T. C., Swapp, R. E. and Kaplan, K. J. (2012), 'Whole-slide imaging: routine pathologic diagnosis', *Advances in Anatomic Pathology* **19**(3), pp. 152–159.

Demir, C. and Yener, B. (2005), 'Automated cancer diagnosis based on histopathological images: a systematic survey', *Rensselaer Polytechnic Institute, Tech. Rep* .

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.

Devroye, L. (1996), *A probabilistic theory of pattern recognition*, Springer.

Farjam, R., Soltanian-Zadeh, H., Jafari-Khouzani, K. and Zoroofi, R. A. (2007), 'An image analysis approach for automatic malignancy determination of prostate pathological images', *Cytometry Part B: Clinical Cytometry* **72**(4), pp. 227–240.

Fox, H. (2000), 'Is H&E morphology coming to an end?', *Journal of Clinical Pathology* **53**(1), pp. 38–40.

Gonzalez, R. C., Woods, R. E. and Eddins, S. L. (2009), *Digital image processing using MATLAB*, Vol. 2, Gatesmark Publishing Knoxville.

Gonzalez, R., Woods, R. and Eddins, S. (2004), *Digital image processing using MATLAB*, Prentice Hall Upper Saddle River, NJ.

Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N. and Yener, B. (2009), 'Histopathological image analysis: A review', *IEEE Reviews in Biomedical Engineering* **2**, 147–171.

Gurcan, M. N., Kong, J., Sertel, O., Cambazoglu, B. B., Saltz, J. and Catalyurek, U. (2007), Computerized pathological image analysis for neuroblastoma prognosis, *in* 'AMIA Annual Symposium Proceedings', Vol. 2007, American Medical Informatics Association, p. pp. 304.

Gurcan, M. N., Pan, T., Shimada, H. and Saltz, J. (2006), Image analysis for neuroblastoma classification: Segmentation of cell nuclei, *in* '28th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, 2006. EMBS'06.', IEEE, pp. 4844–4847.

Harder, N., Mora-Bermúdez, F., Godinez, W. J., Ellenberg, J., Eils, R. and Rohr, K. (2006), Automated analysis of the mitotic phases of human cells in 3d fluorescence microscopy image sequences, *in* 'Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006', Springer, pp. 840–848.

Heckbert, P. (1982), 'Color image quantization for frame buffer display', *SIGGRAPH Comput. Graph.* **16**(3), 297–307.

Ho, S., Bullitt, E. and Gerig, G. (2002), Level-set evolution with region competition: automatic 3-d segmentation of brain tumors, *in* '16th International Conference on Pattern Recognition', IEEE, pp. 532–535.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics* **6**(2), pp. 65–70.

Hu, M., Ping, X. and Ding, Y. (2004), Automated cell nucleus segmentation using improved snake, *in* '2004 International Conference on Image Processing, 2004. ICIP'04.', Vol. 4, IEEE, pp. 2737–2740.

Huang, K., Cooper, L., Sharma, A. and Pan, T. (2006), Fast automatic registration algorithm for large microscopy images, *in* 'Life Science Systems and Applications Workshop, 2006. IEEE/NLM', IEEE, pp. pp. 1–2.

Hughes, M., Marsden, H. B. and Palmer, M. K. (1974), 'Histologic patterns of neuroblastoma related to prognosis and clinical staging', *Cancer* **34**(5), pp. 1706–1711.

Ihlow, A., Held, C., Rothaug, C., Dach, C., Wittenberg, T. and Steckhan, D. (2011), Evaluation of expectation maximization for the segmentation of cervical cell nuclei, *in* 'Bildverarbeitung für die Medizin 2011', Springer, pp. 139–143.

Jafari-Khouzani, K. and Soltanian-Zadeh, H. (2003), 'Multiwavelet grading of pathological images of prostate', *IEEE Transactions on Biomedical Engineering* **50**(6), pp. 697–704.

Joshi, V. V., Cantor, A. B., Altshuler, G., Larkin, E. W., Neill, J. S., Shuster, J. J., Holbrook, C. T., Hayes, F., Nitschke, R., Duncan, M. H. et al. (1992), 'Age-linked prognostic categorization based on a new histologic grading system

of neuroblastomas. a clinicopathologic study of 211 cases from the pediatric oncology group', *Cancer* **69**(8), pp. 2197–2211.

Kass, M., Witkin, A. and Terzopoulos, D. (1988), 'Snakes: Active contour models', *Journal of Computer Vision* **1**(4), pp. 321–331.

Khoury, M., Burnett, L., Mackay, M. A. et al. (1996), 'Error rates in Australian chemical pathology laboratories', *Medical Journal of Australia* **165**(3), pp. 128–129.

Kittler, J. and Illingworth, J. (1986), 'Minimum error thresholding', *Pattern Recognition* **19**(1), pp. 41–47.

Kong, H., Gurcan, M. and Belkacem-Boussaid, K. (2011), 'Partitioning histopathological images: An integrated framework for color-texture segmentation and cell splitting', *IEEE Transactions on Medical Imaging* **30**(9), pp. 1661–1677.

Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H. and Gurcan, M. N. (2009), 'Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation', *Pattern Recognition* **42**(6), pp. 1080–1092.

Kovalev, V. A., Grigoriev, A. Y., Ahn, H.-S. and Myshkin, N. K. (1996), Segmentation technique of complex image scene for an automatic blood cell counting system, *in* 'Proc. SPIE', Vol. 2710, pp. 805–810.

Kumar, S., Ong, S. H., Ranganath, S., Ong, T. C. and Chew, F. T. (2006), 'A rule-based approach for robust clump splitting', *Pattern Recognition* **39**(6), pp. 1088–1098.

Lam, L., Lee, S.-W. and Suen, C. Y. (1992), 'Thinning methodologies-a comprehensive survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(9), pp. 869–885.

Lee, K.-M. and Street, W. N. (2003), 'An adaptive resource-allocating network for automated detection, segmentation, and classification of breast cancer nuclei topic area: image processing and recognition', *IEEE Transactions on Neural Networks* **14**(3), pp. 680–687.

Lehmann, E. L. and Casella, G. (1998), *Theory of point estimation*, Vol. 31, Springer.

Li, C., Xu, C., Gui, C. and Fox, M. (2005), Level set evolution without re-initialization: A new variational formulation, *in* 'Computer Vision and Pattern Recognition (CVPR)', Vol. 1, IEEE, pp. 430–436.

Li, C., Xu, C., Gui, C. and Fox, M. D. (2010), 'Distance regularized level set evolution and its application to image segmentation', *IEEE Transactions on Image Processing* (12), pp. 3243–3254.

Li, K., Lu, Z., Liu, W. and Yin, J. (2012), 'Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake', *Pattern Recognition* **45**(4), pp. 1255–1264.

Lu, Z., Carneiro, G. and Bradley, A. P. (2013), Automated nucleus and cytoplasm segmentation of overlapping cervical cells, *in* 'Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013', Springer, pp. 452–460.

Madhloom, H., Kareem, S., Ariffin, H., Zaidan, A., Alanazi, H. and Zaidan, B. (2010), 'An automated white blood cell nucleus localization and segmentation

using image arithmetic and automatic threshold', *Journal of Applied Sciences* **10**, pp. 959–966.

Matas, J., Chum, O., Urban, M. and Pajdla, T. (2004), 'Robust wide-baseline stereo from maximally stable extremal regions', *Image and Vision Computing* **22**(10), pp. 761–767.

Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M. and Tomaszewski, J. (2008), Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology, *in* 'IEEE International Symposium on Biomedical Imaging. ISBI 2008.', pp. 284–287.

Naik, S., Doyle, S., Feldman, M., Tomaszewski, J. and Madabhushi, A. (2007), Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information, *in* 'Microscopic Image Analysis with Applicationns in Biology Workshop'.

Nam, D., Mantell, J., Bull, D., Verkade, P. and Achim, A. (2012), Active contour based segmentation for insulin granule cores in electron micrographs of beta islet cells, *in* 'Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)', pp. 5339–5342.

Nock, R. and Nielsen, F. (2006), 'On weighting clustering', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(8), pp. 1223–1235.

Otsu, N. (1975), 'A threshold selection method from gray-level histograms', *Automatica* **11**, pp. 285–296.

Park, J. et al. (2010), 'Neuroblastoma: biology, prognosis, and treatment', *Hematology Clinics of North America* **24**(1), pp. 65–86.

Powers, D. (2011), 'Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation', *Journal of Machine Learning Technologies* **2**(1), pp. 37–63.

Qi, X., Xing, F., Foran, D. J. and Yang, L. (2012), 'Robust segmentation of overlapping cells in histopathology using parallel seed detection and repulsive level set', *IEEE Transactions on Biomedical Engineering* **59**(3), pp. 754–765.

Reddick, W. E., Glass, J. O., Cook, E. N., Elkin, T. D. and Deaton, R. J. (1997), 'Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks', *IEEE Transactions on Medical Imaging* **16**(6), pp. 911–918.

Roerdink, J. and Meijster, A. (2000), 'The watershed transform: Definitions, algorithms and parallelization strategies', *Fundamenta Informaticae* **41**(1-2), pp. 187–228.

Rosai, J., Rosai, J. and Ackerman, L. V. (2004), *Rosai and Ackerman's surgical pathology*, 9 ed edn. Previous ed.: published as Ackerman's surgical pathology, 1996.

Rousson, M. and Paragios, N. (2002), Shape priors for level set representations, *in* 'Computer Vision 2002', Springer, pp. 78–92.

Ruiz, A., Kong, J., Ujaldon, M., Boyer, K., Saltz, J. and Gurcan, M. (2008), Pathological image segmentation for neuroblastoma using the GPU, *in* '5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008', IEEE, pp. 296–299.

Russell, R. A., Adams, N. M., Stephens, D. A., Batty, E., Jensen, K. and

Freemont, P. S. (2009), 'Segmentation of fluorescence microscopy images for quantitative analysis of cell nuclear architecture', *Biophysical Journal* **96**(8), pp. 3379–3389.

Sadeghian, F., Seman, Z., Ramli, A., Kahar, B. and Saripan, M. (2009), 'A framework for white blood cell segmentation in microscopic blood images using digital image processing', *Biological Procedures* **11**(1), pp. 196–206.

Sertel, O., Kong, J., Catalyurek, U. V., Lozanski, G., Saltz, J. H. and Gurcan, M. N. (2009), 'Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading', *Journal of Signal Processing Systems* **55**(1-3), pp. 169–183.

Shimada, H., Ambros, I., Dehner, L., Hata, J., Joshi, V. and Roald, B. (1999), 'Terminology and morphologic criteria of neuroblastic tumors', *Cancer* **86**(2), pp. 349–363.

Shimada, H., Umehara, S., Monobe, Y., Hachitanda, Y., Nakagawa, A., Goto, S., Gerbing, R., Stram, D., Lukens, J. and Matthay, K. (2001), 'International neuroblastoma pathology classification for prognostic evaluation of patients with peripheral neuroblastic tumors', *Cancer* **92**(9), pp. 2451–2461.

Soille, P. (2003), *Morphological image analysis: Principles and Applications*, Springer-Verlag, New York.

Ta, V.-T., Lézoray, O., Elmoataz, A. and Schüpp, S. (2009), 'Graph-based tools for microscopic cellular image segmentation', *Pattern Recognition* **42**(6), pp. 1113–1125.

Teot, L. A., Sposto, R., Khayat, A., Qualman, S., Reaman, G. and Parham, D.

(2007), 'The problems of central pathology review: development of a standardized procedure for the children's oncology group', *Pediatric and Developmental Pathology* **10**(3), pp. 199–207.

Tomaszewski, J. E., Bear, H. D., Connally, J. A., Epstein, J. I., Feldman, M., Foucar, K., Layfield, L., LiVolsi, V., Sirota, R. L., Stoler, M. H. et al. (2000), 'Consensus conference on second opinions in diagnostic anatomic pathology who, what, and when', *American Journal of Clinical Pathology* **114**(3), pp. 329–335.

Troxel, D. B. (2004), 'Error in surgical pathology', *The American Journal of Surgical Pathology* **28**(8), pp. 1092–1095.

Tsai, S.-H., Hsieh, Y.-H. and Chen, C.-S. (2013), A novel clustering approach for the segmentation of pathological cells images, *in* '2013 International Conference on Advanced Robotics and Intelligent Systems (ARIS)', IEEE, pp. 49–53.

Turban, E., Sharda, R., Delen, D. and Efraim, T. (2007), *Decision support and business intelligence systems*, Pearson Education.

Valenstein, P. N. and Sirota, R. L. (2004), 'Identification errors in pathology and laboratory medicine', *Clinics in Laboratory Medicine* **24**(4), pp. 979–996.

Wang, H. et al. (2012), 'Clump splitting via bottleneck detection and shape classification', *Pattern Recognition* **45**(7), pp. 2780–2787.

Weyn, B., van de Wouwer, G., van Daele, A., Scheunders, P., van Dyck, D., van Marck, E. and Jacob, W. (1998), 'Automated breast tumor diagnosis and grading based on wavelet chromatin texture description', *Cytometry* **33**(1), pp. 32–40.

Wu, H.-S., Xu, R., Harpaz, N., Burstein, D. and Gil, J. (2005), 'Segmentation of intestinal gland images with iterative region growing', *Journal of Microscopy* **220**(3), pp. 190–204.

Xu, C. and Prince, J. (1997), Gradient vector flow: A new external force for snakes, *in* 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 66–71.

Yang, X., Li, H. and Zhou, X. (2006), 'Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy', *IEEE Transactions on Circuits and Systems I: Regular Papers* **53**(11), pp. 2405–2414.

Yeo, T., Jin, X., Ong, S., Sinniah, R. et al. (1994), 'Clump splitting through concavity analysis', *Pattern Recognition Letters* **15**(10), pp. 1013–1018.

Zack, G., Rogers, W. and Latt, S. (1977), 'Automatic measurement of sister chromatid exchange frequency.', *Journal of Histochemistry & Cytochemistry* **25**(7), pp. 741–753.

Zahn, C. T. et al. (1972), 'Fourier descriptors for plane closed curves', *IEEE Transactions on Computers* **100**(3), pp. 269–281.

Zeng, Z., Strange, H., Han, C. and Zwiggelaar, R. (2013), Unsupervised cell nuclei segmentation based on morphology and adaptive active contour modelling, *in* 'Image Analysis and Recognition', Springer, pp. 605–612.

Zhou, X. and Li, F. (2009), 'A novel cell segmentation method and cell phase identification using markov model', *IEEE Transactions on Information Technology in Biomedicine* **13**(2), pp. 152–157.