

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	A study towards device-independent histological quantification and automatic classification system
著者(和文)	Abdul Aziz Maulana
Author(English)	Maulana Abdul Aziz
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10992号, 授与年月日:2018年9月20日, 学位の種別:課程博士, 審査員:山口 雅浩,小林 隆夫,熊澤 逸夫,金子 寛彦,渡辺 義浩
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10992号, Conferred date:2018/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis



TOKYO INSTITUTE OF TECHNOLOGY

Imaging Science & Engineering Laboratory

4259 Nagatsuda, Midori-ku, Yokohama 226-8503, Japan

Doctoral Thesis

A study towards device-independent histological quantification and automatic classification system

Tokyo Institute of Technology
Interdisciplinary Graduate School of Science and Engineering
Department of Information Processing

Maulana Abdul Aziz

Supervisor:
Professor Masahiro Yamaguchi

Abstract

Digital pathology, especially with recent development of whole slide imaging (WSI) scanners, has great potential in expanding pathological analysis level through fully quantitative pathology. Yet, the variety of imaging system implemented on WSI scanners may affect the WSI quality, and further the quantitative pathology. The main objective of this study was to explore the possibility in developing device-independent histological quantification and automatic classification system. Here, several approaches and new features having potential to improve classification rates were evaluated. Furthermore, several digital correction methods on color and sharpness variety existed in WSIs were investigated. The results showed that the classification rate can be improved by utilizing different set of features and digital corrections are necessary to improve the reliability of the quantification and classification process.

Acknowledgments

All praise is due to Allah, the Creator, the Guardian, and the All-Knowing.

This work has its roots in the teaching, help, patience, inspiration, and support of a great number of people, to whom I wish to express my heartfelt gratitude.

Firstly, I would like to express my sincere gratitude to my advisor Professor Masahiro Yamaguchi for his continuous support, advice, and guidance throughout my study.

I would like to thank the thesis committee; Professor Itsuo Kumazawa, Professor Takao Kobayashi, Professor Hirohiko Kaneko, and Associate Professor Yoshihiro Watanabe.

My deep gratitude to the Indonesia Endowment Fund for Education (LPDP) for their full support on this study.

My sincere thanks to Dr. Tomoya Nakamura, Dr. Yuri Murakami, Dr. Fumikazu Kimura, Mr. Hiroshi Kanazawa, Ms. Kumiko Shiga, and Ms. Minori Okawa for their great assistance in the lab.

My sincere thanks also to Dr. Masahiro Ishikawa, Prof. Naoki Kobayashi, Dr. Tokiya Abe, Prof. Akinori Hashiguchi, Prof. Michiie Sakamoto, Mr. Tomoharu Kiyuna, Ms. Yoshiko Yamashita, and Mr. Akira Saito for their insightful comments and assistance.

I thank my fellow lab mates for their kind assistance and the members of Tokyo Tech Indonesian Student Association for their support.

Last but not least, a special thank you to my parents, my wife (you can do it, *Insha Allah*) and other family members whose love, faith, and constant support fuel all of my endeavors. May Allah SWT grant you the best reward for all of your kindnesses, assistance, and supports.

Table of Contents

Abstract	iii
Acknowledgments	v
Table of Contents	vii
List of Figures.....	xi
List of Tables	xvii
Chapter 1 Introduction	1
1.1. Background	1
1.2. Objectives of the study	6
1.3. Thesis outline	8
Chapter 2 Digital pathology and quantitative pathology.....	9
2.1. Pathological diagnosis.....	9
2.2. Digital pathology, whole slide scanner, and whole slide image	11
2.2.1. Digital pathology	11
2.2.2. Whole slide scanner	11
2.2.3. Whole slide image.....	14
2.3. Quantitative pathology and computer-assisted diagnosis system	18
Chapter 3 Enhancement of automatic classification system of histopathological images	21
3.1. Introduction to hepatocellular carcinoma	21

3.1.1. Liver cancer and HCC	21
3.1.2. Grades of HCC	23
3.2. Liver histological components	25
3.2.1. Structure of hepatic histological specimen.....	25
3.2.2. Segmentation of liver histological components.....	26
3.3. Histological features of liver tissue	28
3.3.1. Nuclear features	28
3.3.2. Trabecular features	28
3.3.3. Tissue changes features.....	29
3.4. Method to enhance automatic classification of HCC images	32
3.4.1. Enhancement of automatic classification of HCC images.....	32
3.4.2. HCC image collection.....	32
3.4.3. HCC image classification.....	33
3.4.4. Results and evaluation on HCC image classification	36
3.5. HCC image classification based on tissue types.....	43
3.5.1. Types of HCC tissue histological pattern	43
3.5.2. Overview of classification system based on tissue type.....	45
3.5.3. Texture-based tissue features of liver	46
3.5.4. Liver image classification based on tissue type.....	49
3.5.5. HCC image classification post tissue classification.....	53
3.6. Discussion	55
Chapter 4 Color correction and its effect on quantitative pathology	57
4.1. Color correction on histopathological images	57
4.1.1. Device-based color correction	57
4.1.2. Stain-based color correction	58
4.2. Evaluation of stain-based color correction on histopathological images acquired by different WSSs.....	59
4.2.1. Experiment workflow.....	59

4.2.2. Effect of stain-based color correction on tissue components	62
4.2.3. Effect of stain-based color correction on nuclei detection.....	66
4.2.4. Effect of stain-based color correction on quantified features	69
4.3 Effect of device-based color correction on histopathological images acquired by different WSSs	78
4.3.1. Experiment workflow.....	78
4.3.2. Effect of device-based color correction on digitized color chart slide	80
4.3.3. Effect of stain-based color correction on tissue components	85
4.4. Effect of combination of device- and stain-based color correction on histopathological images acquired by different WSSs.....	89
4.4.1. Experiment workflow.....	89
4.4.2. Effect of combination of device- and stain-based color correction on nuclei detection	90
4.4.3. Effect of combination of device- and stain-based color correction on quantified features	92
4.4.4. Effect of combination of device- and stain-based color correction on intensity profiles.....	96
4.5. Discussions.....	102
Chapter 5 Sharpness correction and its effect on quantitative pathology.....	103
5.1. Sharpness on histopathological images acquired by different WSSs	103
5.2. MTF-based correction on histopathological images	105
5.2.1. Introduction to Wiener filter	105
5.2.2. Histopathological-images-derived MTF filter	106
5.2.3. Evaluation on MTF-based correction on histopathological images	108
5.3. Gaussian model sharpness correction on histopathological images acquired by different WSSs	112
5.3.1 Sharpness correction based on Gaussian model point spread function	112
5.3.2. Experiment workflow.....	112
5.3.3. Effect of sharpness correction on nuclei detection	114

5.3.4. Effect of sharpness correction on intensity profiles	117
5.3.5. Effect of sharpness correction on quantified features.....	123
5.4. Discussion	134
Chapter 6 Summary and future work.....	135
6.1. Summary	135
6.1.1. Summary of the proposed methods.....	135
6.1.2. Overall concept of device-independent histological quantification and automatic classification system.....	136
6.2. Future Work.....	138
Publications and Conferences.....	139
Bibliography.....	141
Appendices	145
Appendix A.....	145

List of Figures

Figure 1.1 Proportion of global deaths under the age 70 years, by cause of death, in 2012 as reported by WHO [2].	2
Figure 1.2 Most common cancer incidences on each country for male (top) and female (bottom) measured with age-standardized rates indicator and expressed per 100,000 [3].	3
Figure 1.3 Most common cancer mortalities on each country for male (top) and female (bottom) measured with age-standardized rates indicator and expressed per 100,000 [3].	4
Figure 2.1 Different approaches of scanning and focus strategy; (a) tiling with focus every field, (b) tiling with focus nth field, and (c) line scanning with focus map.	13
Figure 2.2 Two WSI of liver tissues [17]. Tissue on the left is a surgical resected sample where the specimen is extracted through operation while tissue on the right is a biopsy resected sample which is extracted by using needles.....	14
Figure 2.3 A digitized liver tissue displayed by a WSI viewer software called Panoramic Viewer developed by 3DHISTECH Ltd.....	15
Figure 2.4 Four digitized tissue specimens displayed side-by-side with NDP.view2 viewing software, which supporting images from NanoZoomer series of Hamamatsu Photonics K.K.....	15
Figure 3.1 The list of cancer incidence and mortality worldwide. Liver cancer incident was ranked as 5th and 9th for male and female respectively, yet 2 nd and 6 th in the case of mortality [43].	22
Figure 3.2 Maps of incidence (top) and mortality (bottom) of liver cancer based on regions worldwide with estimated age-standardized rates (world) per 100,000 [3].....	23
Figure 3.3 HE-stained histopathology images of various liver tissue types; non-tumor liver, well differentiated (G1), moderately differentiated (G2), poorly differentiated (G3) and undifferentiated (G4) [8].	24
Figure 3.4 Structures in a hepatic histological specimen [33].	25

Figure 3.5 ROI images from surgical resected sample (left) and biopsy resected sample (right) [17].	27
Figure 3.6 ROI images before (left) and after (right) the masking process. Nuclei are overlaid with black color mask while other components outside trabecular are overlaid with white color mask [17].	27
Figure 3.7 Overall workflow of the experiment on automatic classification of HCC images system enhancement.	35
Figure 3.8 ROC curves for HCC classification on experiment BB01, CB01, CB02, and CB08 [17].	40
Figure 3.9 ROC curves for HCC classification on experiment SS01, CS01, CS02, and CS08 [17].	41
Figure 3.10 Various types of HCC tissue histologic pattern; (a) Trabecular (or sinusoidal) pattern, (b) Pseudoglandular (or acinar, adenoid) pattern, (c) Sheet (or solid, compact) pattern, (d) Scirrhou pattern, (e) Clear cell. The cytoplasm becomes clear due to fat, glycogen or water, (f) Trabecular pattern containing area with pseudoglandular pattern.	44
Figure 3.11 Diagram of ROI images classification based on the tissue type.	45
Figure 3.12 Four offsets used in extracting the GLCM properties.	46
Figure 3.13 Absorption of hematoxylin and eosin stain in samples. The peak absorptions of hematoxylin and eosin are at 590 nm and 530 nm respectively. These are around the wavelength of green color (546.1 nm) compared to blue (435.8 nm) and red (700 nm) [34], [54].	47
Figure 3.14 Process in generating image for texture features extraction.	48
Figure 3.15 Sub-ROI with different resolutions provides information on the image texture. In the above images, three different regions are highlighted; small area (red), long but thin area (green) and large area (blue). The small area is “gone” in low resolution image while large area can still be seen although with blur effect in the edge.	49
Figure 3.16 Filtering sub-ROIs. Images having too much white area are removed from the data. Image (a), (b), and (c) are covered with 61.90%, 59.58% and 90.59% of white mask respectively. Image (a) and (c) were removed from the classification process.	50
Figure 3.17 Classification results on two ROI images.	52

Figure 3.18 ROI classified as cord-like tissue type by pathologist is classified as trabecular tissue type by the tissue classification.	52
Figure 3.19 Diagram of HCC classification based on the tissue types.	53
Figure 3.20 The result of HCC classification on ROI images that have been categorized between trabecular and non-trabecular type. Cases A and D are the proposed method, while case B and C are shown for comparing the result.	54
Figure 3.21 Effect of masking on distribution difference of nuclear feature.	56
Figure 4.1 The workflows of the device- and stain-based color corrections introduced in [55] and [51].	58
Figure 4.2 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs [58].	59
Figure 4.3 Workflow of experiment on stain-based color correction. 13 ROI images extracted from digitized of three tissue specimens used for the experiment, where each specimen was digitized thrice by three different WSSs [58].	61
Figure 4.4 The effect of stain-based color correction on the three sets of 13 ROI image [58].	63
Figure 4.5 The projection of cluster plots of six tissue components (nuclei, cytoplasm, sinusoid, lymphocyte, fiber, red blood cell) on the a-b chromaticity axes of CIELAB color space [58].	64
Figure 4.6 The number of nuclei detected on each ROI images, both on non-color corrected ROI images (above) and color corrected ROI images (below). In most cases, the number of detected nuclei improved by the application of stain-based color correction regardless of the WSS used to digitize the specimens [58].	67
Figure 4.7 Comparison of nuclear detection results by the prototype system with manual detections (image B). (a) Manual detection. (b)-(d) Before color correction, (e)-(g) after color correction. (b) and (e) WSS-A, (c) and (f) WSS-B, and (d) and (g) WSS-C. The detected nuclei are overlaid with green color [58],	68
Figure 4.8 The 50-percentile values of 12 nuclear features on each case [58].	74
Figure 4.9 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs [58].	77
Figure 4.10 Digitized color chart slide by (a) WSS-A, (b) WSS-B, and (c) WSS-C.	78

Figure 4.11 Obtaining white color on from the digitized color chart slide. A region of 100×100 pixels on scanned white area between the patches was selected to represent the white color for the color correction matrix.....	79
Figure 4.12 Workflow of experiment on device-based color correction. A color chart slide and three WSSs were used to derive color correction matrices B and C. The color correction matrices were then used to correct the color on images captured by WSS-B and WSS-C. Three ROI images of 3 liver specimens were selected for this experiment. Image of the color chart slide is taken from [55].	80
Figure 4.13 The effect of device-based color correction on the ten color patches from the digitized color chart slide. The color difference on the ROI images were reduced by color correction.	81
Figure 4.14 The projection of cluster plots of mean RGB values on each color patch on the L-a chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.	82
Figure 4.15 The projection of cluster plots of mean RGB values on each color patch on the L-b chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.	83
Figure 4.16 The projection of cluster plots of mean RGB values on each color patch on the a-b chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.	84
Figure 4.17 The effect of device-based color correction on the three sets of three ROI images. The colors on the color-corrected images changes although the colors have yet to become similar to the reference images.....	85
Figure 4.18 The projection of cluster plots of six tissue components (nuclei, cytoplasm, sinusoid, lymphocyte, fiber, red blood cell) on the a-b chromaticity axes of CIELAB color space. The number 1 and 2 after the WSS type refer to uncorrected and color corrected version of the WSI.....	86
Figure 4.19 Workflow of experiment on combination of device- and stain-based color correction. Two ROI images from two liver specimens were used, where each specimen was scanned by WSS-A, WSS-B, and WSS-C. Device- and stain-based color corrections were performed sequentially on the ROI images.	90
Figure 4.20 Comparison on the number of nuclei detected on image before and after sharpness correction.	91

Figure 4.21 The 50-percentile values of 12 nuclear features on each case.....	93
Figure 4.22 Intensity images of a nucleus from Case A-1 acquired by three different WSSs and the effect of color corrections on the images.....	97
Figure 4.23 Intensity images of a nucleus from Case E-1 acquired by three different WSSs and the effect of color corrections on the images.....	98
Figure 4.24 The comparison of intensity profiles of the selected nuclei.....	100
Figure 5.1 Effect of image quality on quantitative pathology. (a) In case A, the two nuclei are only detected on ROI image from WSI-A (right) while in case B, the nucleus is only detected on ROI image from WSI -C (left). (b) Both in case E and F, the nuclei are detected on both ROI images. However, the masking area in case F are different between the two images. Image on the left side is from WSI-C while image on the right side is from WSI-A.....	104
Figure 5.2 Image degradation and restoration model.....	105
Figure 5.3 Workflow to derive MTF-based correction on histopathological images, where the filter is based on a reference histopathological image.....	106
Figure 5.4 Workflow to derive MTF-based correction on histopathological images, where the filter is based on a reference histopathological image.....	107
Figure 5.5 Illustration on the process of averaging the magnitude of the image's Fourier transform for each angle in order to generate 1D MTF.	108
Figure 5.6 ROI images used for the experiment.....	109
Figure 5.7 The 2D MTF of each WSS, the 2D MTF-based filters, and the 1D MTF differences between WSS-B and WSS-C with WSS-A.....	109
Figure 5.8 Several ROI images after MTF -based correction.....	110
Figure 5.9 Comparison on the number of nuclei detected on WSIs before and after MTF-based correction.....	111
Figure 5.10 WSIs of a tissue scanned by a WSS at different time, showing different shape of objects due to difference depth of field during scanning process.	111
Figure 5.11 The results of image deblurring with Gaussian model PSFs having several σ values.	113
Figure 5.12 Workflow of experiment on sharpness correction. Two ROI images from two liver specimens were used, where each specimen was scanned by WSS-A,	

WSS-B, and WSS-C. Sharpness correction is implemented ahead of the stain-based color corrections on the ROI images.	114
Figure 5.13 Comparison on the number of nuclei detected on image before and after sharpness correction.	115
Figure 5.14 Image correction on image G.	116
Figure 5.15 Image correction on image K.	116
Figure 5.16 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs.	118
Figure 5.17 Comparison on the intensity values of nucleus B. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are [0,0.0039] and [0,0.0170], respectively....	119
Figure 5.18 Comparison on the intensity values of nucleus C. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are [0,0.0118] and [0,0.0065], respectively....	120
Figure 5.19 Comparison on the intensity values of nucleus E. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are [0,0.0157] and [0,0.0105], respectively....	121
Figure 5.20 Comparison on the intensity values of nucleus F. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are [0,0.0052] and [0,0.0039], respectively....	122
Figure 5.21 The result of nuclei quantification for each nuclear case.	124
Figure 5.22 The 50-percentile values of three nuclei features on each case.	131
Figure 6.1 The concept of device-independent histological quantification and automatic classification system.	137

List of Tables

Table 3.1 The list of nuclear, trabecular, and tissue changes features in this study. For each feature introduced by Kiyuna et al., six statistical distributions are calculated, which are 10, 30, 50, 70, 90 percentiles and standard deviation. In total, 99 features were extracted, consisting 78 nuclear features, 10 trabecular features, and 11 tissue changes features.....	30
Table 3.2 Details of data used for the experiments.	33
Table 3.3 Experiment results using various set of features with biopsy resected samples as the training data.....	37
Table 3.4 Experiment results using various set of features with surgical resected samples as the training data.....	38
Table 3.5 Experiment results using various set of features with a mix of surgical and biopsy resected samples as the training data.	39
Table 3.6 Details of condition between the method by Kiyuna et al. [14] and the proposed method, in the case of surgical resected samples as training data.	42
Table 3.7 Comparison on true positive rates between Kiyuna et al. [14] method and the proposed method, in the case of surgical resected samples as training data.....	42
Table 3.8 Classification results on sub-ROI images based on the pattern types.....	51
Table 3.9 Classification results on ROI images based on the pattern types.	51
Table 4.1 Comparison between the results of nuclei detection by the prototype system and manual detection [58].	69
Table 4.2 Value difference of 12 quantified nuclear features without and with color correction. Each feature is represented by its 50-percentile value [58].....	70
Table 4.3 Effect of device-based color correction on the tissue components.	88
Table 4.4 A set of five images for each case in the experiment to evaluate the combination of device- and stain-based color correction.	89
Table 4.5 Differences on the number of nuclei detected between images before and after sharpness correction.	92

Table 4.6 Differences on the nuclear feature extracted from stained-based color-corrected image A without and with device-based color correction where the nuclear feature on reference image (WSI-A).	95
Table 4.7 Differences on the nuclear feature extracted from stained-based color-corrected image E without and with device-based color correction where the nuclear feature on reference image (WSI-A).	95
Table 4.8 Objects used in the experiment.	96
Table 4.9 The sub-cases on each object.	97
Table 4.10 Description on sub-figures in Figure 4.24.	99
Table 5.1 Differences on the number of nuclei detected between images before and after sharpness correction.	115
Table 5.2 Differences on the feature of nuclear B on images without and with sharpness correction where the feature from WSI-A becomes the reference.	128
Table 5.3 Differences on the feature of nuclear C on images without and with sharpness correction where the feature from WSI-A becomes the reference.	128
Table 5.4 Differences on the feature of nuclear E on images without and with sharpness correction where the feature from WSI-A becomes the reference.	129
Table 5.5 Differences on the feature of nuclear F on images without and with sharpness correction where the feature from WSI-A becomes the reference.	129
Table 5.6 Differences on the nuclear feature extracted from image G before and after sharpness correction with the nuclear feature on reference image (WSI-A).	133
Table 5.7 Differences on the nuclear feature extracted from image K before and after sharpness correction with the nuclear feature on reference image (WSI-A).	133

Chapter 1

Introduction

1.1. Background

Noncommunicable disease (NCD) is a non-infectious disease. It is often labeled as chronic diseases, which progress slowly over an extended period due to a combination of genetic, physiological environmental and behavioral factors. According to Global Health Observatory (GHO) data by World Health Organization (WHO) which provides statistics related to health, the number of global deaths in 2015 was 56.4 million and 39.5 million (70%) of the cases were due to NCDs. The four highest causes of NCD deaths were from cardiovascular diseases (17.7 million deaths), cancer (8.8 million deaths), respiratory diseases (3.9 million deaths), and diabetes (1.6 million deaths) [1]. These numbers are higher compared to the statistics in 2012, where 38 million out of a total of 56 million deaths were caused by NCDs as reported in the Global Status Report on Noncommunicable Diseases 2014 by WHO [2]. About 75% of mortality due to NCDs occurred in low- and middle- income countries and populations, having 42% (in 2012) and 48% (in 2015) of the cases occurred to the patients having age before 70 years in these regions. Figure 1.1 shows the proportion of global deaths under the age of 70 years, by cause of death in 2012 as reported in [2]. Four primary major risk factors that lead to the rise of deaths by NCDs are the use of tobacco, lack of physical inactivity, alcohol use, and unhealthy diets (e.g., lack of fruit and vegetable intake). On a broader perspective, NCD cases are not only affecting the individual's quality of life but may also cause domino effects on the families, communities and further the socioeconomic impacts in the society.

Cancer is a common word for an enormous group of diseases which may affect any portion of the body. It is also recognized as malignant tumor or neoplasm. For a cancer patient, abnormal cells in his/her body are developed in a vast speed and grow exceeding their normal boundaries, invading the surrounding parts of the body and further spread to other organs. The process of spreading is called metastases, and this leads to the primary cause of death in cancer cases. Cancer may affect people from many countries. Nevertheless, common cancer types in a region may differ to the types in another region as cancer causes are often related to its risk factors. Figure 1.2 and Figure 1.3 exhibit the

most common cancer incidences and mortalities in each country, respectively. Cancer's risk factors can be divided into two main categories, i.e. behavioral-dietary risks (e.g., high body mass index and the four primary major risk factors of NCDs) and carcinogenic infections, which include physical carcinogen (e.g., ultraviolet), chemical carcinogen (e.g., asbestos), and biological carcinogen (e.g., viruses).

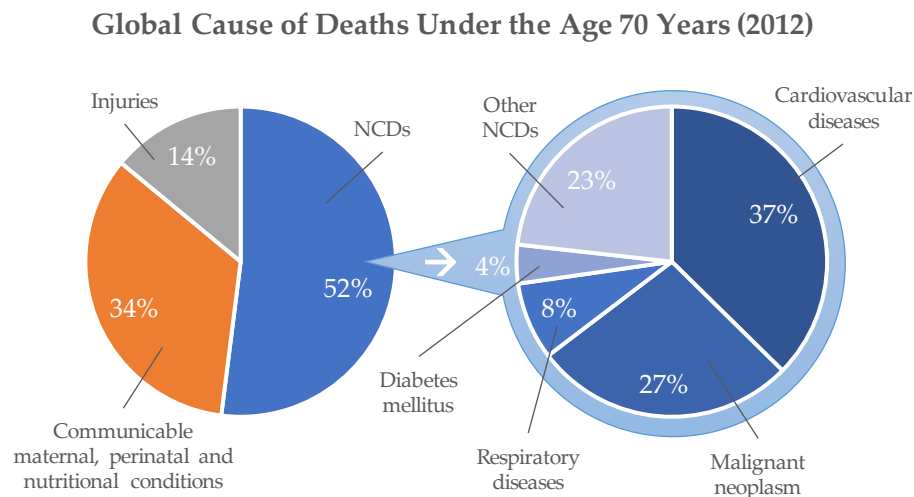


Figure 1.1 Proportion of global deaths under the age 70 years, by cause of death, in 2012 as reported by WHO [2].

Cancer causes about one out of six deaths worldwide, which leads cancer to become the second highest cause of death. In 2012, 32.6 million people were living with cancer within five years of diagnosis worldwide, where 14.1 million were new cases within the year, as reported by International Agency for Research on Cancer (IARC), a specialized cancer agency of WHO. Studies by IARC shows that the cancer incidences have been increasing and the trend is projected to continue. WHO estimated that the number of new cases would keep increasing, where the number of cancer cases will close to 25 million in the 2030s [2]. Although high-income countries have the highest incidence of cancer, highest cancer mortality occurred in low- and middle-income countries due to late-stage diagnosis and poor clinical outcomes. These facts reflect the importance of providing best services for cancer detection, diagnosis, and treatment to improve survival rates worldwide.

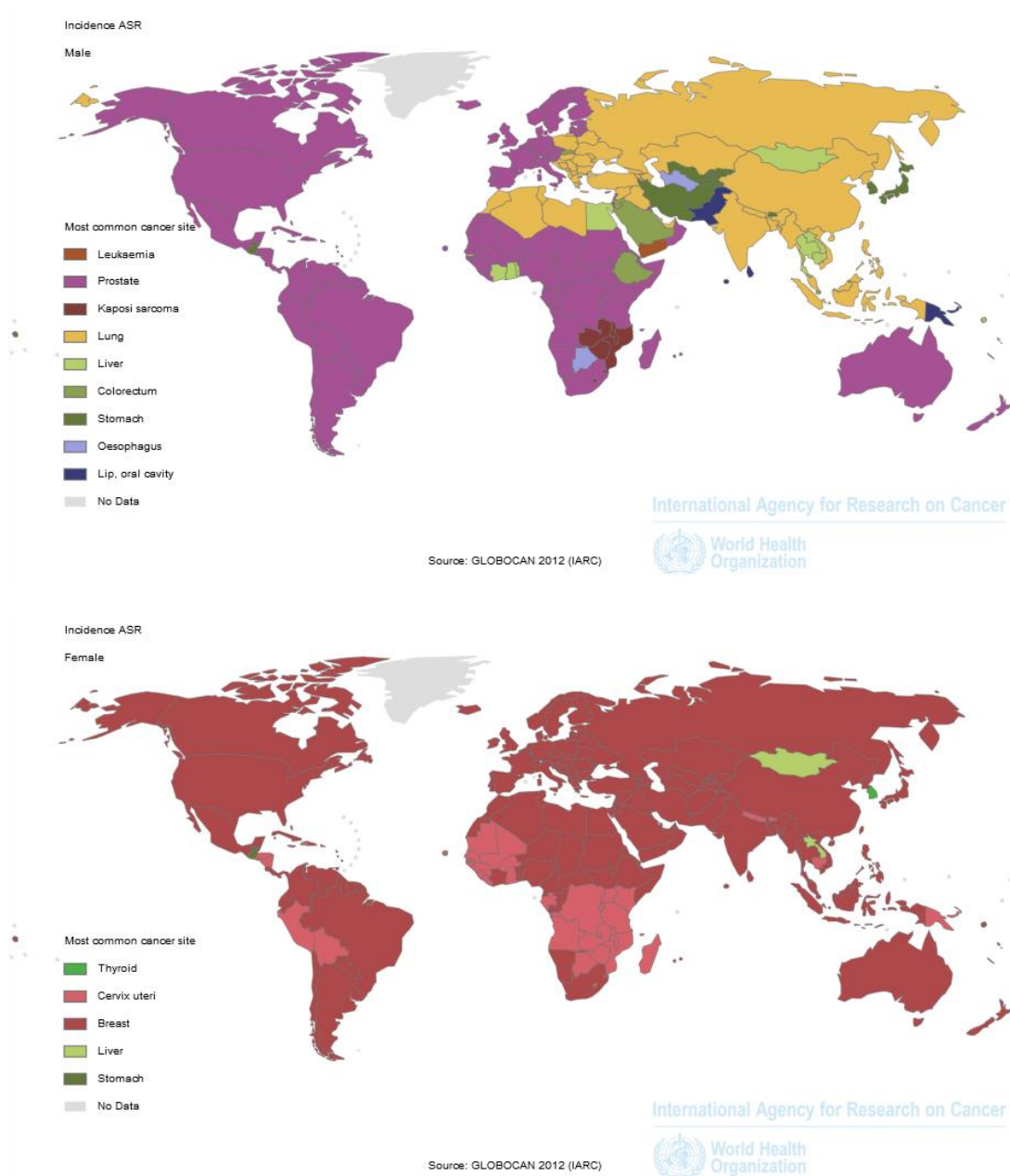


Figure 1.2 Most common cancer incidences on each country for male (top) and female (bottom) measured with age-standardized rates indicator and expressed per 100,000 [3].

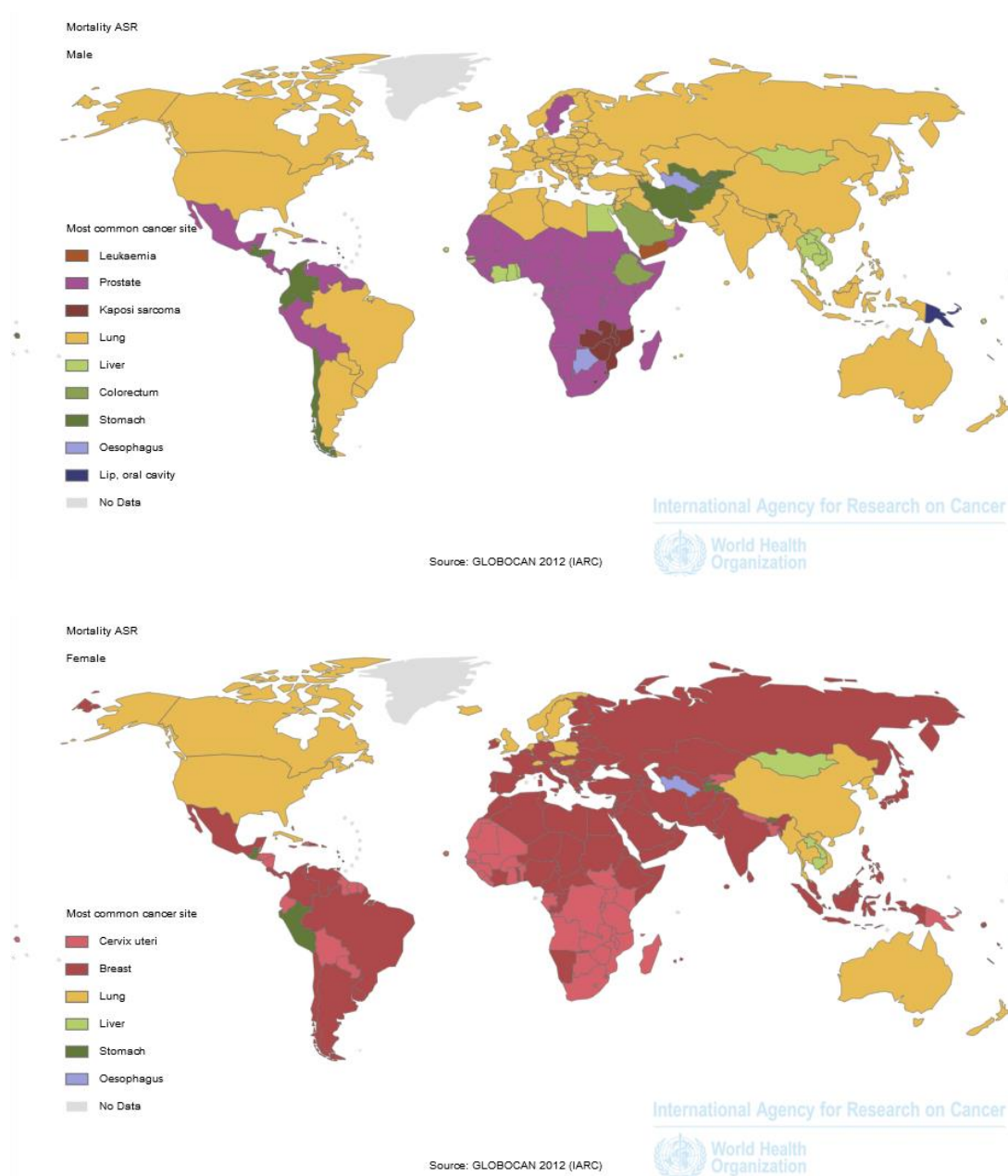


Figure 1.3 Most common cancer mortalities on each country for male (top) and female (bottom) measured with age-standardized rates indicator and expressed per 100,000 [3].

Histopathological analysis plays important role on cancer detection. In the conventional process, tissue resected from the patient, either through surgery or biopsy are fixed, stained, and put into glass slides so that pathologist can analyze the specimens by microscope and identify the disease and determine its grades, stages, and prognosis. Nevertheless, with the increasing number of cancer cases, small ratio of number of pathologists compared to the population, and lack of access to pathological diagnosis for those who live in country sides, it becomes necessary to assist the pathologist in order to expedite and improve the diagnosis.

On the other hand, recent breakthrough in computational power and digital microscopy, in addition to the advancement of computer vision techniques and image analysis algorithms, have prompted the application of such technologies on histopathological image analysis in various type of cancers, including lung cancer [4], breast cancer [5], [6], prostate cancer [7], thyroid cancer [4], and also liver cancer [8], [9]. With the development of whole slide scanner (WSS), it is possible to digitize the pathological samples so that the monitor-based examination is performed with a color monitor on the digitized specimens. Furthermore, objective quantification of tissue morphology will be obtained by applying image analysis on the histopathological images, which often labeled as whole slide image (WSI). Various works on digital pathology have been performed, with the hope to aid the pathologists and improve the diagnosis process, such as the development monitor-based diagnosis [10], feature quantification [11]–[16], and computer-aided diagnosis (CAD) system [8], [12], [17]–[19]. The CAD system will hopefully expedite the process of histopathological image analysis and relieve the workload of pathologists. Pathologists can focus on cases that are more difficult to analyze as the CAD system filters the benign (non-malignant) tissues. Furthermore, the system can also improve the quantitative analysis of the pathology images by providing quantitative data of the tissues, and thus reduce the subjectivity of pathologist opinions. In fact, more information can be obtained by applying image analysis techniques on digitized specimens so that pathologists can perform higher level diagnosis.

Aside all possible improvements that can be offered by digital pathology in the field of pathological diagnosis, digital pathology is in its early phase. There are still concerns with the utilization of WSI, especially for clinical uses. Even though some reports have mentioned the use of digital pathology as primary diagnosis in Canada and some European regions [20], [21], lack of regulatory on WSS standard becomes one of the drawbacks on further utilization of WSI for clinical uses.

1.2. Objectives of the study

The main goal of this research is to perform a study on the possibility of developing a histological quantification and automatic classification system that is independent to the device used in acquiring the histopathological images. The goal can be elaborated into two major objectives. The first objective is to elaborate the possibility on enhancing automatic classification system of histopathological images by applying tissue structure analysis. The second objective is to develop a digital correction method on the WSI in order to achieve device-independent histopathological quantification system. As the case study, the research will use an anatomical pathology diagnosis assisting system developed by NEC Corporation which is used for automatic classification of histopathological images of liver cancer called hepatocellular carcinoma [17].

Nuclei features have been utilized as the main object on pathological diagnosis by the pathologists. The similar case is also happened in digital pathology as features related to nuclear are usually used as the main objects in digital pathology. Furthermore, nuclei in various organs have similar characteristics and not depends on the tissue type. Therefore, a lot of studies have been carried out on nuclei quantification, thus the methods on nuclei quantification are more advanced and well established compared to the other tissue components. Nevertheless, tissue structure has its own role in pathological diagnosis and some works to segment and quantify tissue structures have been developed in order to include more information for the pathological analysis purpose. However, the effectiveness of these new features on cancer detection based on histopathological images classification have not been evaluated. This work will evaluate the effect of the tissue structure features that have been developed in HCC images classification. Moreover, the importance of image masking that will ignore certain regions during feature quantification is analyzed. Then, to further enhance the system, a method to classify HCC images based on the tissue characteristics will be explored, since some quantified features are not suitable to classify certain type of liver tissue.

On the other hand, the prototype system developed by NEC Corporation is based on a single type WSS. Yet, it is important for the prototype system to be able to process images captured from different devices as there are various type of WSSs available in the market. However, each device utilizes its own imaging system and the variety of the imaging system may lead to differences on image quality, which later on may affect the quantification result. Therefore, it is necessary to develop a digital correction method to

restore histopathological images acquired by different devices through digital correction. Here, several color correction methods on histopathological images are evaluated. The evaluation focuses mainly on its effect on quantified features. Lastly, sharpness correction methods on histopathological images are introduced and evaluated in order to overcome the focus problems founded on WSIs acquired by different WSSs.

1.3. Thesis outline

This thesis is organized into six chapters.

The first chapter introduces the background of the study, its objectives and the outline of the thesis.

The second chapter provides some backgrounds regarding pathological diagnosis, digital pathology, whole slide scanner, whole slide image, quantitative pathology, and computer-aided diagnosis system.

Chapter 3 explains the methods to enhance the automatic cancer classification of histopathological images by combining nuclear and tissue structural features, combination of digitized samples, image masking, and classification of histopathological images based on tissue type.

Chapter 4 focuses on evaluation of color correction methods on histopathological images and its effect on quantitative pathology.

Chapter 5 describes the evaluation of sharpness corrections on histopathological images and its effect on quantitative pathology.

Chapter 6 provides the summary of the proposed methods, description on the overall concept of the device-independent histological quantification and automatic classification systems, and suggestions for future work.

Chapter 2

Digital pathology and quantitative pathology

This chapter provides some backgrounds regarding pathological diagnosis, digital pathology, whole slide scanner, whole slide image, quantitative pathology, and computer-aided diagnosis system that may be referred in the following chapters.

2.1. Pathological diagnosis

Pathology is one of the major fields in medicine where diseases are observed by pathologists, a physician who has the specialty in interpreting laboratory test and evaluating cells, tissues, and organs. Histopathology is focusing in the study of cells or tissues by using the microscope in order to identify the disease and to determine its grades, stages, and prognosis.

Pathological diagnosis starts by removing tissue sample called specimen from the patient either through surgery or biopsy needle. The pathologist then performs a macroscopic or gross examination by the naked eye on the specimen that has been preserved and labeled, and make description on the visual features, size, and procedures that have been done to the specimen. Some part of the specimen may be prepared for some other tests, such as molecular, gene, or protein tests, based on requirement. The specimen that is going to be examined with a microscope needs to be prepared as a slide. For this, the specimen will be fixed by using formalin in order to prevent decay and to preserve the form. The fixed specimen is then cut into a single layer of cells (called as histologic sections) so that it is transparent enough to be observed by microscope. After that, the specimen is stained to highlight the features of interest. Hematoxylin and eosin (H&E) stain is the most common staining where the cell nuclei are stained purplish blue while cytoplasm and several other extra-cellular components are stained red. The stained specimen is placed on a glass slide before covered with a coverslip [22].

Once the glass specimen is ready, the pathologist studies the section by using a microscope, whether the tissue contains tumor, and if it exists, either it is benign (non-cancerous tumor) or malignant (cancerous and have the possibility to spread to the other organs). Based on his or her observation, the pathologist creates a pathology report

describing the types, arrangement, and normality of the cells, and other essential features for diagnosis. The pathologist will also provide a description of the diagnosis based on his or her observation on the biopsy, during the macroscopic examination, processing, and the microscopic examination [23].

Pathological diagnosis plays an essential role in cancer diagnosis and staging, as further treatments on the cancer patients will be based on a report provided by the pathologists. In cancer cases, structures of the tissue will change, forming diverge architecture than normal tissue. Therefore, quantitative measurement on the tissue structure become important in histopathological analysis. The pathologists use their knowledge and experiences in examining and quantifying the tissue components. Even though to become a specialist in pathology, pathologists have taken years of training program, pathologists often use subjective terms in measuring the tissue components. For example, the tissue nuclei are larger compared to in normal tissue. Such method lacks absolute reference. The increment number of cancer patients worldwide has also increased the workload of pathologists. The situation becomes worse when the ratio between pathologists and the population is very small or unevenly distributed [8].

2.2. Digital pathology, whole slide scanner, and whole slide image

2.2.1. Digital pathology

The utilization of digital images in pathology have been used for many decades, starting with telepathology in the late 1960s where a digital image of a partial pathology slide was transmitted to different location [21]. As technology advanced, the application of digital devices on pathology becomes varies. Now, digital pathology may refer to all related technologies that support digital slides in the process of pathological diagnoses, which include management systems on image and laboratory, digital dictation, digital image analysis, electronic labelling and tracking on specimens, and so on.

The implementations of digital technology have been able to improve the safety, quality, and efficiency on the diagnoses workflow. It is reported in [21] that up to 5% errors on patient identification in the pathology laboratory were happened during accession and copying patient's details. By using barcode and installing system that manage the information on the laboratory, errors due to misidentification can be reduced. The management system can also improve the efficiency as it becomes easier to monitor the process of diagnosis, from specimen extraction on the patient, data management (storing, searching, retrieving, sharing, etc.), and up to the creation pathological report. In the aspect of quality, digitizing pathological specimens opens the opportunity of implementing image analysis techniques, such as for feature quantification, digital measurement, comparison of slides side-by-side, and so on.

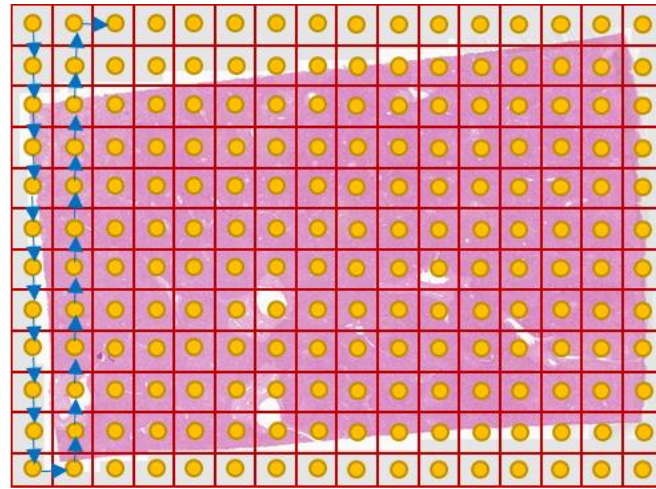
2.2.2. Whole slide scanner

The development of whole slide scanner (WSS) has been one of the factor that expedite digital pathology enhancement. It is a device that can digitize pathological slide in a high magnification, so that the sample can be observed in computer screen. Generally, WSS use lenses and light sources having similar standard to microscope lenses and lightning sources [24]. Therefore, effective resolution of WSS can be similar to the microscope.

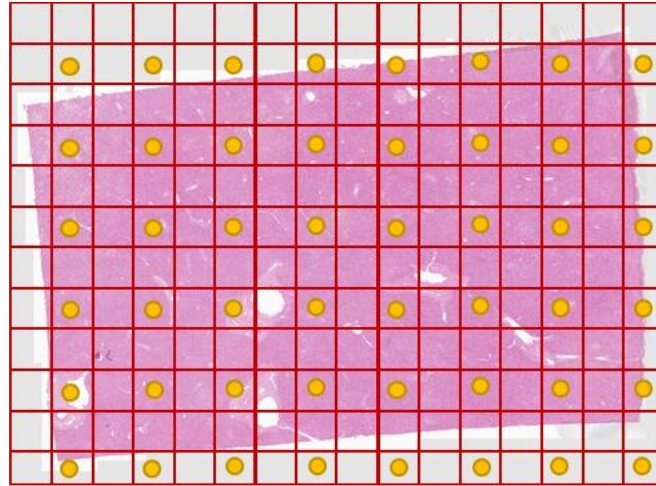
There are numerous types of slide scanner with various technologies and imaging systems on it. Generally, the processes of digitizing tissue slide are similar, yet the techniques applied on each step may vary depends on the vendor. The process to digitize

tissue slide starts with image focusing. Histopathology specimens, which are three-dimensional (3D) objects, are turned into two-dimensional (2D) images at the certain focus. This process may cause information lost from the 3D slide. One way to overcome this problem is by implementing z-stack scanning method on the WSS, where the specimen is digitized into several layers. The focus strategy implemented on the scanner may vary and it is usually associated with the scanning process. Depending on the size of the tissue area that going to be scanned, WSS may acquire hundreds or thousands of individual images. This process is controlled by a robotic system that moves the lens or the slide, and it may produce different type of individual images. In order to produce one virtual slide, these individual images will be compiled through stitching or tiling process. In tiling, the slide is scanned as a series of rectangular tiles having focus strategy either by focus on every field or focus n^{th} field. Meanwhile, line scanning scans the slide as a series of long-narrows strips just like how common image scanner works. Figure 2.1 illustrate the variations on scanning approaches and its focus strategy. Once a single virtual slide has been produced, image processing techniques may be applied on the virtual slide in order to improve its quality such as image sharpening and color and brightness adjustment. In the end, the image may be compressed for practicality since the size of the original virtual slide is very large. Here, various algorithms are utilized for the process.

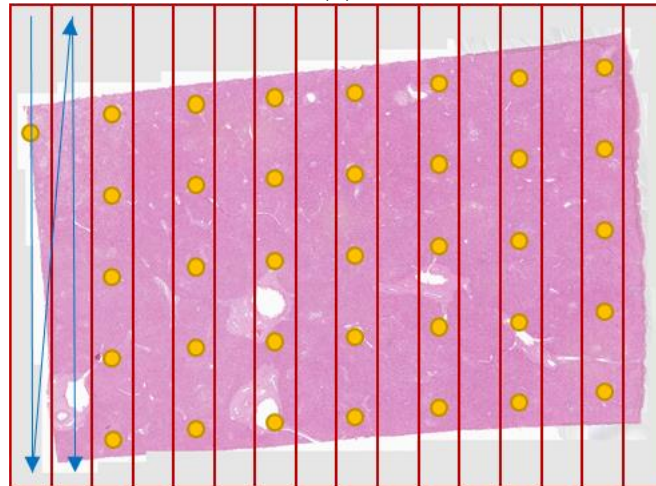
The process of image acquisition can be performed semi-automatically or fully-automatic. An operator will load the digitized images and check the quality of the image. This procedure is necessary in order to confirm the focus quality of the slide and whether the tissue have been captured according to the requirement.



(a)



(b)



(c)

Figure 2.1 Different approaches of scanning and focus strategy; (a) tiling with focus every field, (b) tiling with focus nth field, and (c) line scanning with focus map.

2.2.3. Whole slide image

Whole slide image (WSI) is the digitized tissue sample produced by WSS and also often referred as virtual microscopy. This is because, instead of analyzing cells or tissues in a glass slide by using conventional microscope, the glass slide is digitized by using slide scanners at high magnification so that it can be displayed and analyzed through computer screen. In addition to the pathology field, WSI has been used in education, research, publication, telepathology, and primary diagnosis. Figure 2.2 shows two examples of digitized liver tissues.

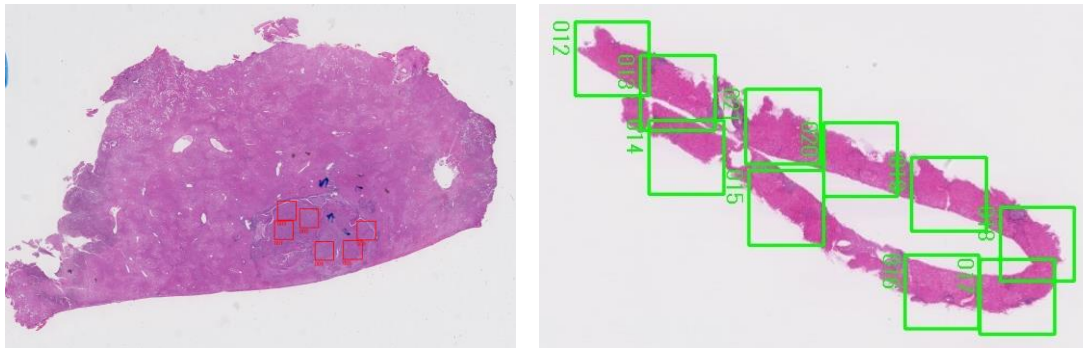


Figure 2.2 Two WSI of liver tissues [17]. Tissue on the left is a surgical resected sample where the specimen is extracted through operation while tissue on the right is a biopsy resected sample which is extracted by using needles.

WSI are displayed in monitor and viewed using specialized software. In most cases, virtual slide vendors will develop their own software for slides acquired by their WSS. Overall, all of the viewing systems have similar basic functions to observe and navigate the digitized tissue, such as small thumbnail of the WSI showing current location, zoom in and out functions using button on the screen, keyboard or other input devices, function snapshot, annotations, measurement, image analysis features, and option to view multiple WSI side-by-side. Figure 2.3 shows an example of WSI viewer while Figure 2.4 shows four WSIs displayed side-by-side by another WIS viewer.

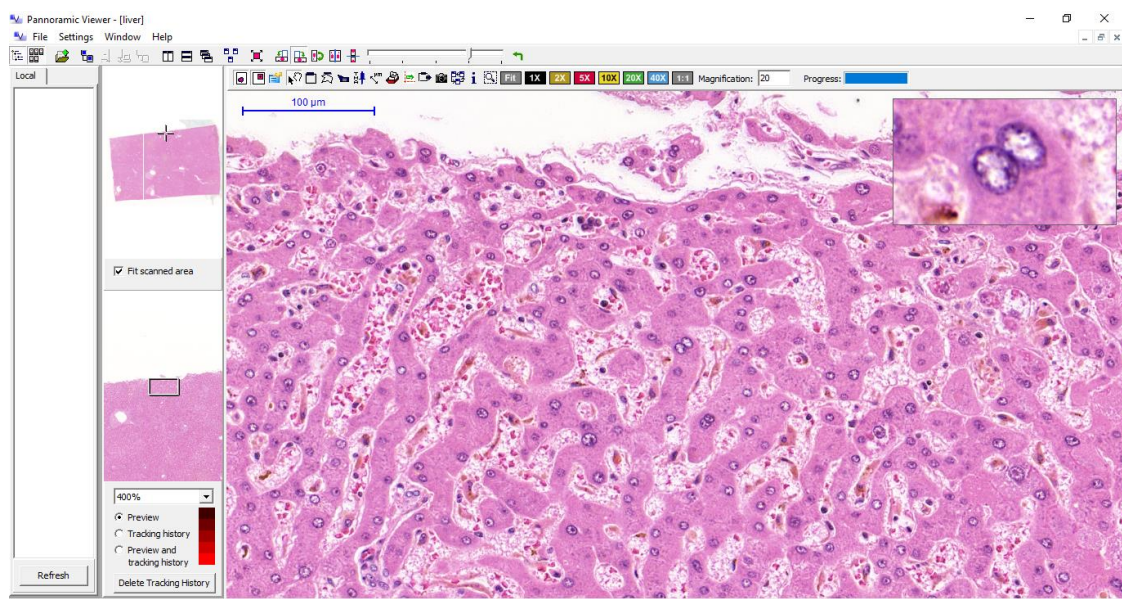


Figure 2.3 A digitized liver tissue displayed by a WSI viewer software called Pannoramic Viewer developed by 3DHISTECH Ltd.

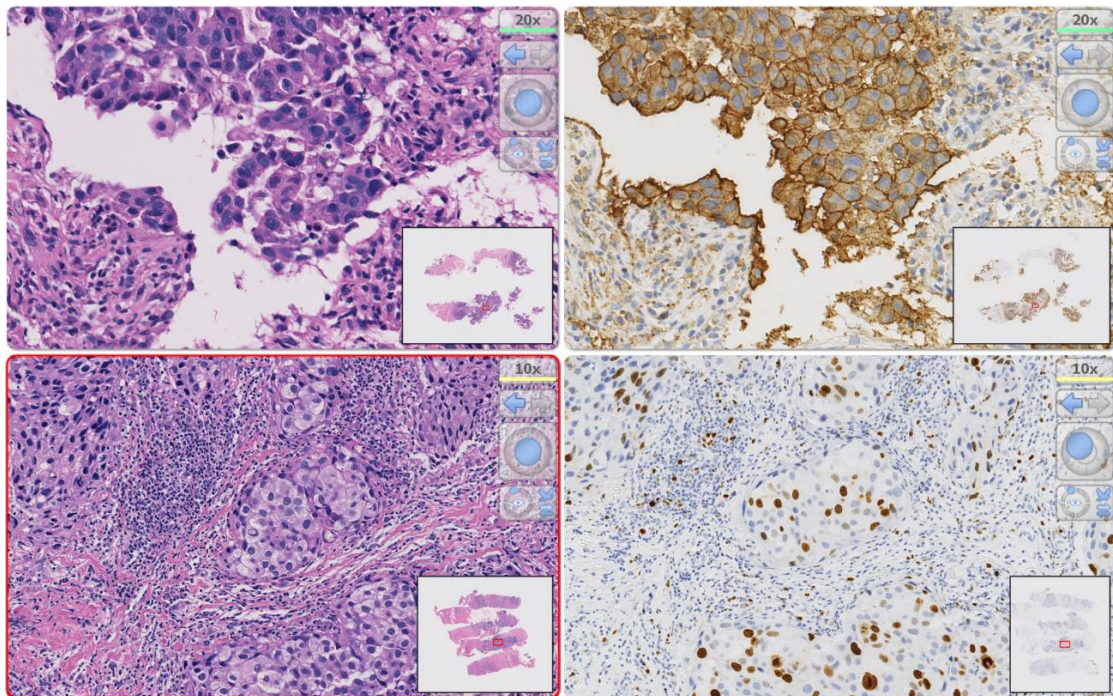


Figure 2.4 Four digitized tissue specimens displayed side-by-side with NDP.view2 viewing software, which supporting images from NanoZoomer series of Hamamatsu Photonics K.K.

Various benefits can be obtained from WSI compared to the conventional glass slide. Some of the advantages are it facilitates external quality assurance, allows teleconferencing and rapid remote consultation, easiness for training and teaching processes, provides better integration to the laboratory management system, easier to be archived and retrieved, and so on. Nevertheless, the WSI implementation may also have several disadvantages such as the pathologists may be more comfortable in observing glass slide and in the aspect of cost since it requires significant IT infrastructure in order to manage the images.

On the other hand, the quality of a WSI plays an essential part in the quantitative analysis of histopathological images. The quality of color on WSI is one of the crucial factors that need to be considered when image processing techniques are going to be implemented on the image. There are several factors that may lead to color variations on histopathological images, which are consisted of staining protocol, scanning device used to digitize the tissue specimen, hardware to display the WSI, and viewing software of the image.

Tissue sample needs to be stained in order to highlight its features so that it is observable. Without staining the sample, it will be difficult to distinguish the tissue components by the human eye due to the lack of contrast in cell organelles. Staining protocol is the steps need to be accomplished during the preparation of a tissue sample, which includes slicing the specimens into thin layers and the staining process. The procedures performed may cause color varieties on the stained tissue. For example, tissue sample that is stained with higher concentration of stain will provides sample with stronger color intensities. On the other hand, tissues having larger thickness will also tend to have stronger color intensities than thinner one.

Color difference on WSIs due to scanner difference is basically caused by the variation of components equipped on the device. It may be caused by the specification of the optical and hardware components (e.g., light source), or the software design of the color reproduction implemented on the scanner. As mentioned in Section 2.3, there is no established regulation which determine the standard of WSS. Therefore, each manufacturer has its own standard in producing the devices which lead to variety of image quality, including the color.

Color variety of the WSI can also be found during the exhibition of the digitized sample. Monitor or display is the device used to show the digitized sample. Color variations when presenting the WSI through the monitor can be affected by the settings and characteristics of the device, such as the contrast, luminance, resolution, color gamut and primaries. On the other hand, the software used to view WSI is generally produced by the same vendors of the WSS. Here, the parameters for WSI color reproduction in the software are usually set according to the characteristics of the WSS. Therefore, displaying a WSI with viewing software that is developed by another vendor may also lead to color difference of the WSI.

2.3. Quantitative pathology and computer-assisted diagnosis system

The quantitative analysis on the morphology of cell and tissue components has a long history. Müller provided first microscopic description of human malignant tumors in 1838, while the first evidence showing that cancer nuclei are larger than normal cell were most likely delivered by Heiberg and Kemp in 1929 [25]. The quantification on tissue components in pathological analysis plays important role in pathological diagnosis since it provides objective distinction between benign and malignant lesions, tumor grading, prediction of prognosis, and response on therapy. Compared to DNA cytometry, quantification based on morphometric techniques for cancer cases are simpler and inexpensive [25].

Quantitative pathology on digital pathology provides measurement with higher precision as numerical representation offer objectivity. It is also believed that quantification on digital pathology have potential in delivering knowledge to understand complex processes of malignant development and progression [26]. Various studies have been carried out in order to quantify digitized tissue components. Nuclear becomes the most object to be studied in quantitative pathology as shown in [13], [14], [16], [27]–[31]. The nuclear features that are usually becomes subject for quantification includes the morphological related nuclei, such as size (area, diameter, perimeter, short and long axes), shape, roundness, ellipticity, and concavity [25], and texture related nuclei as texture can show important characteristic that can be used for object identification in an image [32]. Nevertheless, researches related to tissue structure can also be found as shown by [12], [33]–[39]. In general, the works on quantitative pathology on digital pathology require the implementation of image analysis techniques on the WSI.

Computer-aided diagnosis (CAD) system is a system to assist and support pathologists in analyzing and diagnosing histopathological images. By combining technologies in image analysis and the pathology fields, the system can analyze cells or tissues that have been digitized, including by performing quantitative pathology on the histopathological images, thus minimize variations of opinion during diagnosis process. In fact, more information can be obtained from the CAD system than the conventional pathological method, allowing pathologists to perform higher level diagnosis, such as to analyze the disease process by understanding the biological mechanisms [40]. *e-Pathologist*[®], an anatomical pathology diagnosis system by NEC Corporation, is an

example of CAD systems. The system has been utilized to assist pathologists in analyzing and detecting various types of cancers, including colon cancer, stomach cancer, and breast cancer [8]. And by employing machine learning techniques, the extracted information can be used to classify pathology images, either it is cancer or non-cancer. The system can also relieve the pathologists' workload by screening out obviously benign tissues so that pathologists can focus on more difficult cases. Moreover, by combining the quantitative pathology provided by the system with the other information related to the patient, the pathologist will be able to perform higher-level diagnosis. Recent works related to the CAD system includes the utilization of deep learning on WSI analysis [41], [42].

Chapter 3

Enhancement of automatic classification system of histopathological images

This chapter is focused on the methods to enhance the classification rates of histopathological images through tissue structures segmentation, features selections, and image classification based on the tissue types. Here, images of hepatocellular carcinoma (HCC) are used as the objects of the study. The process and result of the classification are also explained.

3.1. Introduction to hepatocellular carcinoma

3.1.1. Liver cancer and HCC

In 2012, liver cancer was listed as the fifth most common cancer for men and ninth for women as shown in Figure 3.1. Nevertheless, at the same time, liver cancer was ranked as the second most common cause of cancer death worldwide after lung cancer, having nearly 9.1% of the total death (746,000 deaths) [43]. The high rate of mortality in liver cancer is due to its very poor prognosis. The highest incidence rates of liver cancer are in East Asia, Southeast Asia, Northern Africa and Western Africa as shown in Figure 3.2. The most common type of primary liver cancer is hepatocellular carcinoma. Primary cancer means that the tumor is originally from the mentioned organ, which in this case is liver.

HCC is a malignant (cancerous) tumor with hepatocellular differentiation. It is the second most common cancer in Asia and fourth in Africa with men having a considerably higher risk [44]. Chronic infection of hepatitis B virus (HBV) is the primary cause of HCC and followed by hepatitis C virus (HCV) chronic infection [45], [46]. Other causes of HCC include alcohol cirrhosis, tobacco (smoking), and aflatoxin. The cancer is often diagnosed when the survival time is measured in months which leads to high death rates [47]. In 2002, it was recorded 416,882 cases of deaths due to HCC where 94% of the cases occurred in the first year after diagnosis [44]. One of the difficulties in detecting this type of cancer is due to the requirement of specific expertise for the pathologist in diagnosing histopathology images of HCC, as the cancer diagnosis

heavily relies on the result of the pathological analysis. According to a study by Pazgan-Simon et al. in [48], there are no effective treatment methods available, at least up to 2015. Nevertheless, HCC can be successfully treated if it can be diagnosed early enough, although some treatments are only suitable for certain types of HCC.

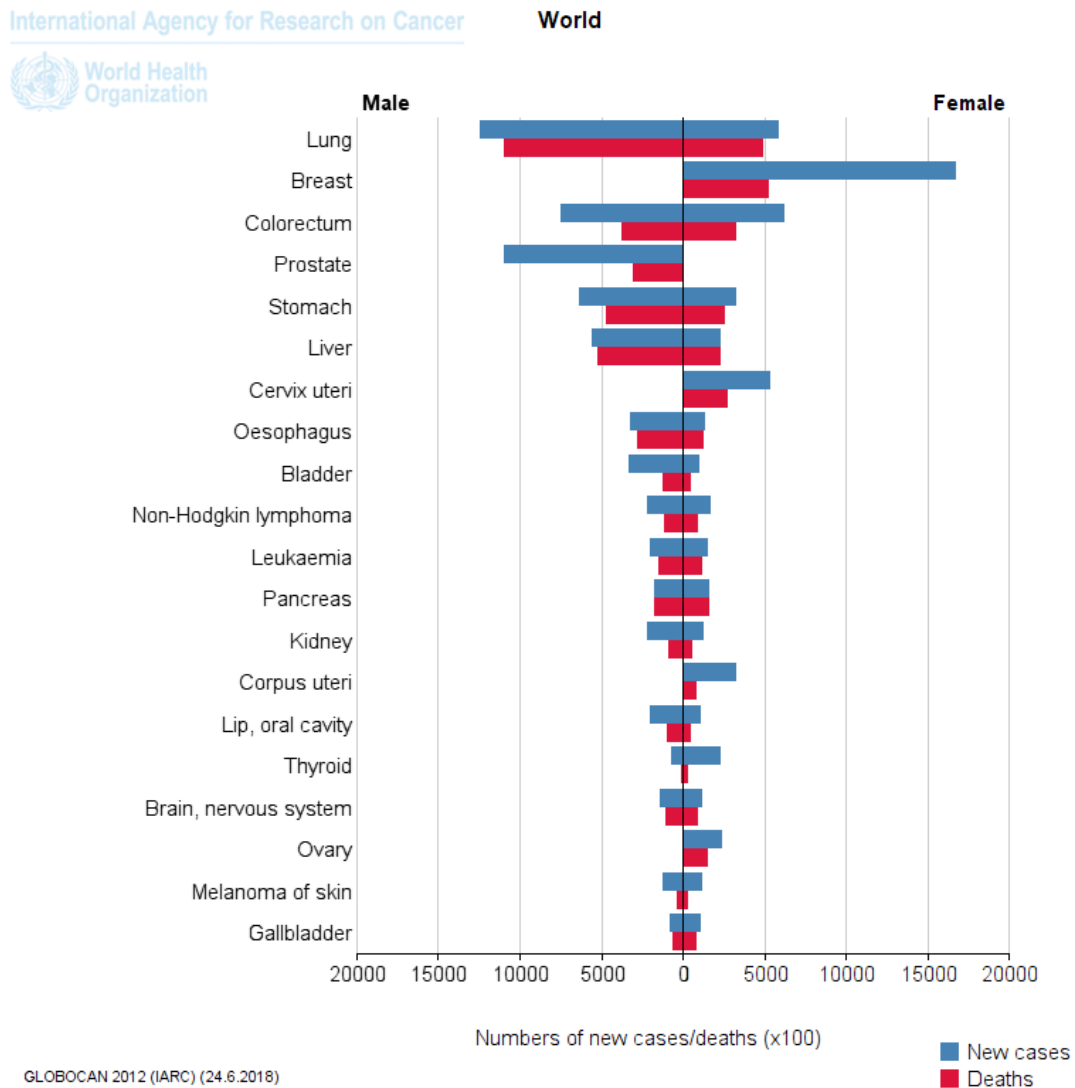


Figure 3.1 The list of cancer incidence and mortality worldwide. Liver cancer incident was ranked as 5th and 9th for male and female respectively, yet 2nd and 6th in the case of mortality [43].

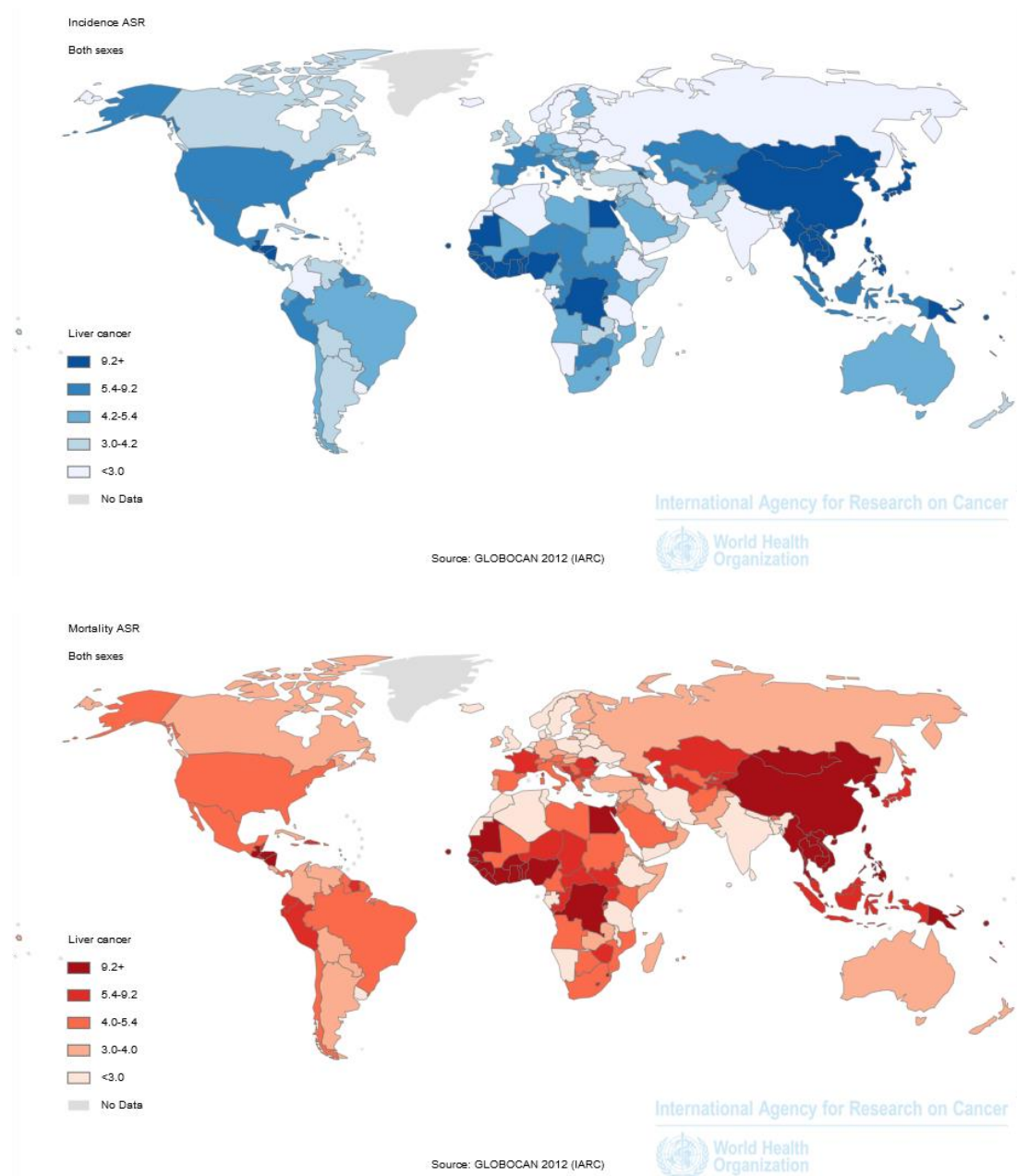


Figure 3.2 Maps of incidence (top) and mortality (bottom) of liver cancer based on regions worldwide with estimated age-standardized rates (world) per 100,000 [3].

3.1.2. Grades of HCC

Histological grading provides valuable prognostic information since the grade describe the biological behavior of the tumor. There are several grading systems of HCC [49]. The Edmondson and Steiner grading system (EGS) introduced by Edmondson and

Steiner is the grading system often utilized by the pathologists for HCC cases. The system divided HCC cases into four grades based on its histological differentiation [50]. Here, the differentiation grades show the similarity of the tumor tissue with its original normal tissue, where the cancer cells in lower grades look more like normal cells where the tumor tend to grow and spread slower. In details, the four grades of HCC in EGS are listed as follow.

- 1) **Grade I (well differentiated).** Mostly small tumor cells arranged in thin trabecular in early-stage tumors. According to WHO, there are 6 characteristic features of the early-stage tumors, consisting (1) increment of cell density (more than twice of the healthy liver) which also increase the nuclei-cytoplasm ratio, (2) growth of thin trabecular pattern irregularly, (3) pseudoglandular structures, (4) fatty change, (5) unpaired arteries, and (6) intratumoral portal tracts [44].
- 2) **Grade II (moderately differentiated).** Here, the cells are larger with abnormal nuclei, and pseudoglandular structures existed. Also, the trabecular become thicker up to three or more cells.
- 3) **Grade III (poorly differentiated).** This grade is rarely found in early-stage tumors. The nuclei are getting larger and more hyperchromatic compared to Grade II. The cytoplasm is granular and acidophilic although less than Grade II.
- 4) **Grade IV (undifferentiated).** Less differentiated having hyperchromatic nuclei, and the trabecular pattern is lost.

Figure 3.3 shows images with various HCC grades along with the healthy tissue type.

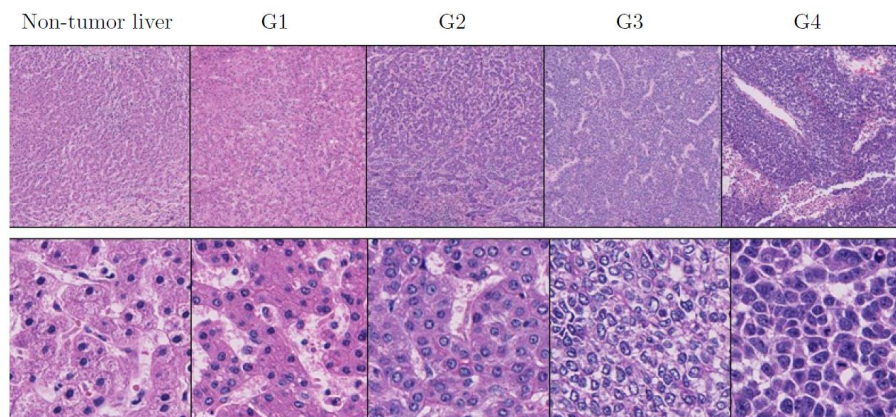


Figure 3.3 HE-stained histopathology images of various liver tissue types; non-tumor liver, well differentiated (G1), moderately differentiated (G2), poorly differentiated (G3) and undifferentiated (G4) [8].

3.2. Liver histological components

3.2.1. Structure of hepatic histological specimen

Hepatic (or liver) is the largest gland of the human body. Hepatic histological specimens are constructed by several components as shown in Figure 3.4. Hepatic cells, known as hepatocytes, are comprised of nuclei and cytoplasm. The cells are arranged in trabeculae and dispersed from the central vein towards the periphery of hepatic lobule. In hematoxylin and eosin (H&E) stained liver tissue, hepatocyte nuclei are round with violet color and located in the central of the hepatocytes. Cytoplasm are the area around the nuclei and died in red or red-purple color. Meanwhile, sinusoids can be seen as the regions separating the hepatic cells having only weakly colored by the H&E stain. Fibers have streamline-form with red or pink color. Lymphocytes are the round shape with dark violet just like nuclei. However, they have smaller size and are located in the sinusoids and fibers regions.

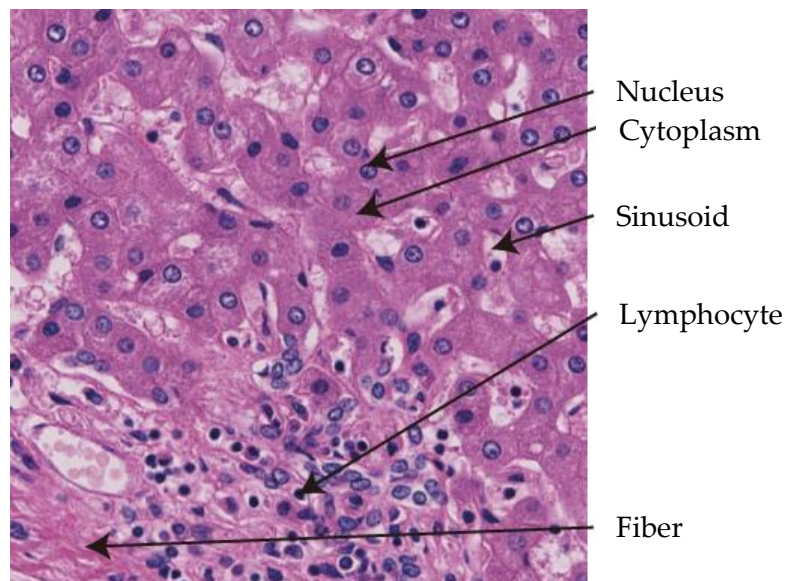


Figure 3.4 Structures in a hepatic histological specimen [33].

In the case of HCC, several changes in the structures occurred. For instance, the nuclei will grow larger and become more heterogeneous. The number of nuclei in a unit area is increased as the cells are multiplied. The cells also often contain fat. Moreover,

while hepatocytes are arranged in one line on the healthy liver tissue, the trabecular structure will become more irregular, and the cells are multiplied to 2-3 lines.

3.2.2. Segmentation of liver histological components

As described in Subsection 3.1.2, liver tissues are constructed by several components. With regards to HCC diagnosis, features from specific components can be more crucial compared to the other. Hepatocyte nuclei, major components of trabecular other than the cytoplasm, have been utilized as one of the main sources of features used by the pathologist in HCC diagnosis. However, liver tissue does not consist of hepatocyte only; thus there are nuclei that belong to the other cells in the tissue. Besides, some components appear relatively similar to hepatocyte nuclei such as lymphocytes. Both nuclei and lymphocytes have round shape and dark color compared to its surrounding in H&E stained image, although lymphocytes have a smaller size. When employing image processing to analyze HCC, it is necessary to identify nuclei that belong to hepatocyte only to get reliable hepatocyte nuclear features.

Several methods to segment the histological components have been developed to overcome the problem, including sinusoids [34], fat droplets, stroma, hepatic cell cords (trabecular), and nuclei [14], [33]. Segmenting liver structures can provide important information that will facilitate recognition on each component. For example, by segmenting and extracting sinusoid regions, hepatic cell cords and stroma regions can be recognized. Additionally, some features can only be derived when the histological components have been segmented first, such as the thicknesses of trabecular [34]. From the developed methods, 6 masked images were generated for each region-of-interest (ROI) images that are, (1) fat mask, (2) fiber mask, (3) glass mask, (4) nuclei mask, (5) red blood cell (RBC) mask, and (6) sinusoid mask. Note that glass mask is mostly used in the case of biopsy samples compared to the surgical samples as the biopsy samples may contain a lot of glass area as shown in Figure 3.5. Meanwhile, Figure 3.6 shows the example of a masked ROI image along with its original image.

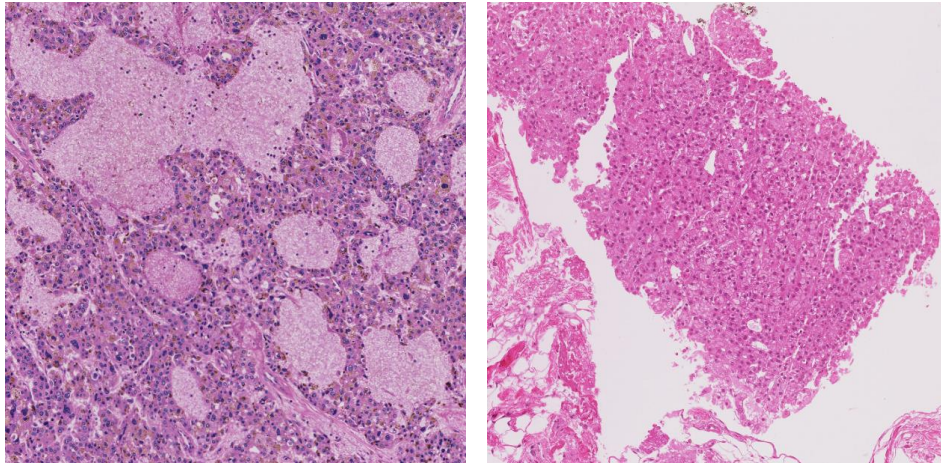


Figure 3.5 ROI images from surgical resected sample (left) and biopsy resected sample (right) [17].

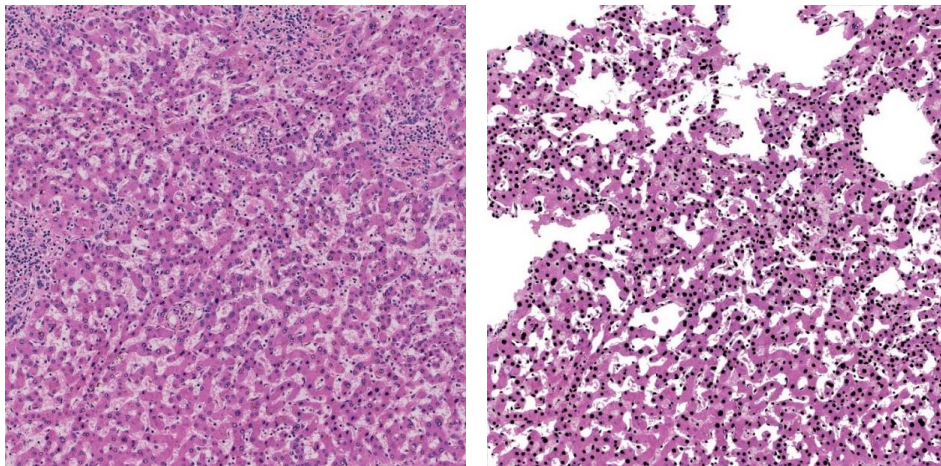


Figure 3.6 ROI images before (left) and after (right) the masking process. Nuclei are overlaid with black color mask while other components outside trabecular are overlaid with white color mask [17].

3.3. Histological features of liver tissue

3.3.1. Nuclear features

In 2013, Kiyuna et al. introduced 11 nuclear features to classify HCC images, along with two structural features automatically. The nuclear features were computed from the nuclei contours and the chromatin textures inside the nuclei. In the process, first the nuclei centroids were determined through edge extraction and a Hough transform on a color map of hematoxylin intensity generated from the ROI image. Initial candidate contours were then obtained by using iterative snake method. The nuclei contours were extracted from candidate contours having similarity with the underlying images above a certain threshold. Meanwhile, the nuclei chromatin textures were characterized by Gray level co-occurrence matrix (GLCM) methods. In the process, the features were obtained by getting the maximum value of the features calculated from four directions ($\theta = 0, \pi/4, \pi/2$, and $3\pi/4$ radian) and three distances ($d = 2, 4, 6$ pixels).

In this study, all of the 11 features related to nuclear were extracted. Six features were extracted from the nuclei contours, consisting (1) area (S), (2) periphery (ℓ), (3) circularity ($C = 4\pi S/\ell$), (4) length of long axis of elliptic fit (a), (5) length of short axis of elliptic fit (b), and (6) ratio of short and long axis lengths ($e = b/a$). Meanwhile, four features were computed from the nuclei chromatin. They are (1) texture angular second moment of GLCM, (2) texture contrast of GLCM, (3) texture homogeneity of GLCM, and (4) texture entropy of GLCM. Another feature, (11) cell contour complexity, was obtained from successive contour lengths at different thresholds. For each feature, six statistical distributions were calculated, which consisted of 10, 30, 50, 70, 90 percentiles and standard deviation [14], producing 66 nuclear related features in total.

Note that the nuclear features utilized by Kiyuna et al. were extracted from histopathological images that have yet to be extracted. In this study, the same types of nuclear features were extracted from images that have been masked by the stroma segmentation first, such that those features are derived from hepatocytes only.

3.3.2. Trabecular features

While the pathologists have been utilizing nuclei as the main features in analyzing histopathology images of HCC, the structure changes of other hepatic tissue components

are also utilized by the pathologists as additional information. Kiyuna et al. in the previous study also introduced two features related to liver structures in addition to the 11 nuclear features, which were nuclei density and trabecular thickness [14]. The nuclei density was calculated based on the number of neighboring nuclei within 30 pixels of diameter while trabecular thickness was calculated based on the number of circles packed between sinusoids. Similar to the nuclear related features, six statistical distributions were also calculated for the nuclei density and trabecular thickness.

Later, to quantify some other characteristics of liver tissue structure, Komagata et al. and Ishikawa et al. developed techniques to quantify some morphological features for hepatic trabecular analysis related to HCC cases [12], [33], [35], [36], [39]. For example, hepatocyte trabecular will get thicker and may become multilayer in the case of cancer. Overall, new ten trabecular features were extracted. They are (1) nuclei-cytoplasm ratio, (2) nuclei density calculated based on the number of nuclei per stroma in a unit area where sinusoids have been removed, (3) average number of layers, (4), standard deviation of number of layers, (5) average of nuclei eccentricity, (6) standard deviation of nuclei eccentricity, (7) graph-based nuclei distance, (8) graph-based nuclei alignment, (9) sinusoid weighted sum of mutual angle distribution histogram (WSMA), and (10) trabecular WSMA.

3.3.3. Tissue changes features

In addition to trabecular features, the number of features was also extended by including 11 features of tissue changes introduced by Ishikawa et al. [33], [37] and Murakami et al. [51]. These are features related to the state of fat droplets, cytoplasm, and stroma. This group of features describes the characteristics of the tissue and not directly related to cancer. Nevertheless, the features will provide additional information to the nuclei and trabecular features for the classification. The features are (1) area ratio of fatty degeneration, (2) number of fat droplets, (3) average value of fat droplet area, (4) standard deviation of fat droplet area, (5) red color of cytoplasm, (6) green color of cytoplasm, (7) blue color of cytoplasm, (8) average value of clearness index, (9) standard deviation of clearness index, (10) area ratio of stroma, and (11) structural feature of stroma.

Table 3.1 below listed the complete features of nuclear, trabecular and tissue changes used in this study.

Table 3.1 The list of nuclear, trabecular, and tissue changes features in this study. For each feature introduced by Kiyuna et al., six statistical distributions are calculated, which are 10, 30, 50, 70, 90 percentiles and standard deviation. In total, 99 features were extracted, consisting 78 nuclear features, 10 trabecular features, and 11 tissue changes features.

No.	Feature	Type
1 – 6	Area (10, 30, 50, 70, 90 percentiles, standard deviation)	Nuclear feature
7 – 12	Periphery (10, 30, 50, 70, 90 percentiles, standard deviation)	
13 – 18	Circularity (10, 30, 50, 70, 90 percentiles, standard deviation)	
19 – 24	Length of long axis of elliptic fit (10, 30, 50, 70, 90 percentiles, standard deviation)	
25 – 30	Length of short axis of elliptic fit (10, 30, 50, 70, 90 percentiles, standard deviation)	
31 – 36	Ratio of short and long axis lengths (10, 30, 50, 70, 90 percentiles, standard deviation)	
37 – 42	Texture angular second moment of GLCM (10, 30, 50, 70, 90 percentiles, standard deviation)	
43 – 48	Texture contrast of GLCM (10, 30, 50, 70, 90 percentiles, standard deviation)	
49 – 54	Texture homogeneity of GLCM (10, 30, 50, 70, 90 percentiles, standard deviation)	
55 – 60	Texture entropy of GLCM (10, 30, 50, 70, 90 percentiles, standard deviation)	
61 – 66	Cell contour complexity (10, 30, 50, 70, 90 percentiles, standard deviation)	
67 – 72	Nuclei density (10, 30, 50, 70, 90 percentiles, standard deviation)	
73 – 78	Trabecular thickness (10, 30, 50, 70, 90 percentiles, standard deviation)	

Table 3.1. (continued).

No.	Feature	Type
79	Nuclei-cytoplasm ratio	Trabecular feature
80	Nuclei density calculated based on the number of nuclei per stroma in a unit area where sinusoids have been removed	
81	Average number of layers	
82	Standard deviation of number of layers	
83	Average of nuclei eccentricity	
84	Standard deviation of nuclei eccentricity	
85	Graph-based nuclei distance	
86	Graph-based nuclei alignment	
87	Sinusoid weighted sum of mutual angle distribution histogram (WSMA)	
88	Trabecular WSMA	
89	Area ratio of fatty degeneration	Tissue changes feature
90	Number of fat droplets	
91	Average value of fat droplet area	
92	Standard deviation of fat droplet area	
93	Red color of cytoplasm	
94	Green color of cytoplasm	
95	Blue color of cytoplasm	
96	Average value of clearness index	
97	Standard deviation of clearness index	
98	Area ratio of stroma	
99	Structural feature of stroma	

3.4. Method to enhance automatic classification of HCC images

3.4.1. Enhancement of automatic classification of HCC images

Kiyuna et al. [14] have introduced an automatic classification of HCC images based on nuclear and structural features, as briefly explained in Subsections 3.3.1 and 3.3.2, which consisted of features 1–78 in Table 3.3. Their experiment was performed on a set of histological images of surgical resected samples. Nevertheless, the common sample in pathological diagnoses is the biopsy specimen which has different characteristic compared to specimens extracted through surgery. The surgical resected sample is large, provides stable features, and may display variations of tissue types. Meanwhile, the biopsy specimen is thin and torn as it is removed by using a needle. Therefore, the inclusion of surgical resected sample as training data on the classification system may have potential to improve the classification rates.

On the other hand, several structural features in addition to features related to tissue changes on liver samples have been developed as described in Subsections 3.3.2–3.3.3. Nonetheless, the effectiveness of these features on HCC image classification has not been evaluated. Moreover, it is important to exclude tissue components that have similar appearance to the hepatocyte nuclear, such as the lymphocyte, from nuclear quantification in order to get reliable nuclear features. This can be carried out by utilizing the tissue segmentation described in Subsection 3.2.2.

The following subsections focus on improving the performance of automatic classification of HCC images. Three approaches are proposed; (1) using a mix of surgical resected samples and biopsy samples as the training data, (2) inclusion of structural and tissue changes features on the classification process, and (3) removing non-hepatocyte nuclear during nuclear quantification. A series of experiments are performed and the effects of each approach on the classification performance are evaluated.

3.4.2. HCC image collection

The experiments were performed on a collection of histopathology images supplied by the Department of Pathology at Keio University School of Medicine, Tokyo, Japan. The images consisted of 147 and 116 H&E stained whole slide images (WSI) of surgical and biopsy resected samples respectively. The surgical samples were collected from 111

patients, varying from 1 to 16 samples for each patient. Meanwhile, the biopsy samples were collected from 108 patients, having 1 or 2 samples per patient. From the surgical samples, 1076 ROI images were extracted as 2048×2048 pixels of TIFF images (2–11 images per slide), where 543 of the images were annotated as HCC positive. As for the biopsy samples, 1054 ROI images having the same size and image type were extracted (1–27 images per slide), where 550 images were HCC positive. Table 3.2 shows the details of the data used in the experiments.

Table 3.2 Details of data used for the experiments.

	Surgical samples	Biopsy samples
Number of patients	111	108
Number of samples	147	116
Number of ROI images	1076	1054
• Non-cancer cases	553	504
• Cancer cases	543	550

All slide images were captured by using the NanoZoomer 2.0HT slide scanner (Hamamatsu Photonics K.K., Hamamatsu, Japan) at 20x magnification, which is equivalent to 0.46 $\mu\text{m}/\text{pixel}$. The produced WSIs are in the format of NanoZoomer Digital Pathology Image before being converted to TIFF images. These samples had been labeled by experienced pathologists at the department. Additionally, the pathologist had also graded the HCC positive cases from the surgical samples based on EGS. In order to reduce the color variation due to the staining process, color correction processes introduced by Murakami et al. [51] were implemented on the histopathological images before extracting the features from the images.

3.4.3. HCC image classification

Before implementing the classification on the HCC images, data preprocessing was implemented on the quantified features. There are several methods to preprocess data for machine learning purpose, such as rescale data, standardize data, normalize data, and binarize data. Each data preprocessing method may produce better benefits in

accordance to the algorithm used for the classification process. In this work, rescale data was implemented, where each feature in the training data was linearly scaled between [0,1]. The test data would then be rescaled by the equation

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the original value of the feature, while x_{min} and x_{max} are the minimum and maximum values for the features in the training data respectively. Rescaling the data may prevent attributes having smaller numeric ranges being dominated by greater numeric, since the scaled data will have relatively equal numeric range. For example, the range of values on some nuclear features, such as circularity and texture ASM of GLCM, are only between [0,1], while features like area, contour complexity, and texture contrast of GLCM may reach hundreds. Rescaling these features will produce features having numeric ranges around [0,1]. Another benefit of rescaling the data is it will avoid numerical difficulties during the calculation process.

In the process of classifying the HCC images, a supervised learning model known as support vector machine (SVM) was employed. A library developed by Chang et al. [52] called LibSVM was utilized for the experiments. The library provides four types of kernel for the classification process, which are linear, polynomial, radial basis function (RBF) and sigmoid. In this work, RBF was selected for the kernel type. The kernel can handle both linear and non-linear relation between the class labels and the attributes, since linear kernel is a special case of RBF. RBF kernel also has less hyperparameters than polynomial kernel and has fewer numerical difficulties. To find the best parameters for the RBF kernel, five-fold cross-validation. More detailed explanation on the library along with its tool can be access in [53].

To evaluate the classification performance, five-fold cross validation was once again used. When partitioning the ROI images, ROIs from the same slide are kept on the same group to avoid ROIs from the same slide being divided into both training and test data. In the process, some combinations are made on the sets of training data from both biopsy and surgery samples. The experiments were also performed on several sets of features to evaluate the new features. As described in the previous subsection, four groups of features were available as listed below.

- 78 unmasked nuclear and structural features (as introduced by Kiyuna et al.)
- 78 masked nuclear and structural features
- 10 trabecular features
- 11 tissue changes features

To see the effect of masking, some experiments on the 78 unmasked nuclear features and 78 masked nuclear features were performed so that the rates of each classification could be compared. The masked nuclear features were then combined with ten trabecular and 11 tissue changes features to evaluate the effect of the new features. Figure 3.7 describes the overall workflow of the process.

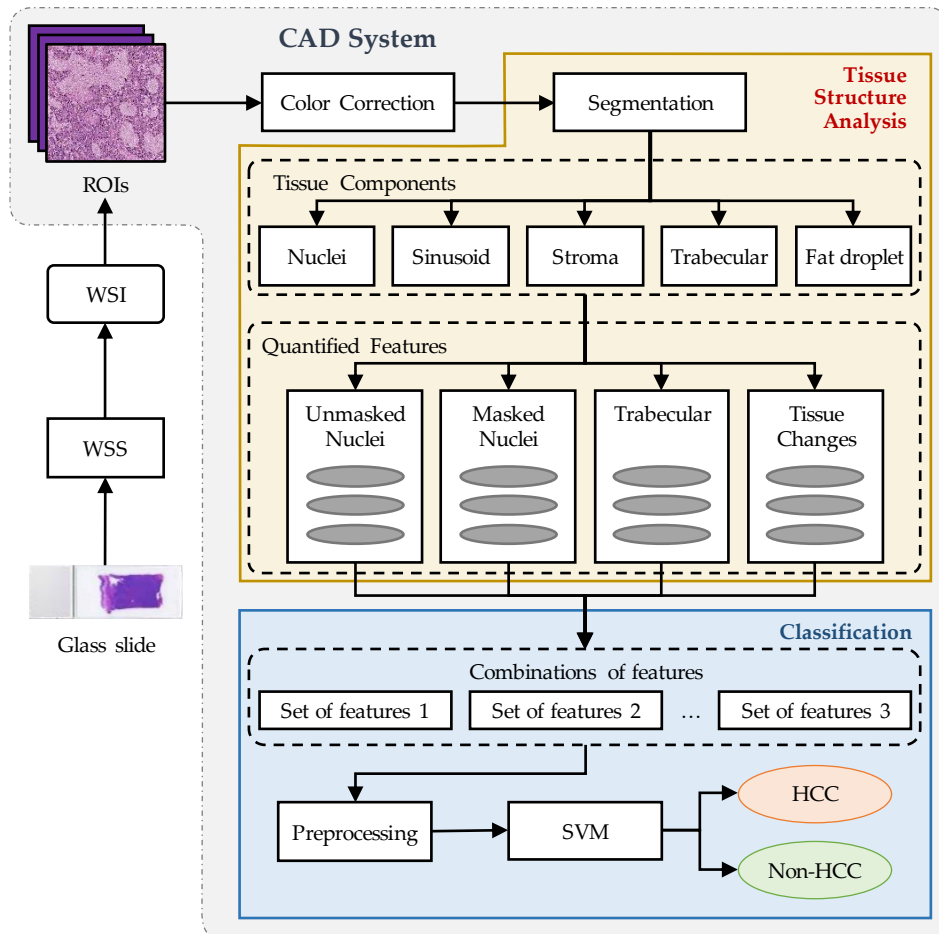


Figure 3.7 Overall workflow of the experiment on automatic classification of HCC images system enhancement.

3.4.4. Results and evaluation on HCC image classification

To evaluate the effectiveness of the segmentation on tissue components and the extracted features on the HCC images classification, classification based on eight sets of feature combinations were compared. The eight sets of features are listed below.

- (1) 78 unmasked nuclear and structural features
- (2) 78 masked nuclear and structural features
- (3) 72 unmasked nuclear and structural features minus trabecular thickness
- (4) 72 masked nuclear and structural features minus trabecular thickness
- (5) 10 trabecular features
- (6) 11 tissue changes features
- (7) 72 unmasked nuclear and structural features, 10 trabecular, and 11 tissue changes features
- (8) 72 masked nuclear and structural features, 10 trabecular, and 11 tissue changes features

The reason on removal of the “trabecular thickness” feature introduced by Kiyuna et al. was because features on the number of layers (both average and standard deviation) within the trabecular features are more comprehensive in representing the liver structural characteristics in terms of the trabecular thickness than the earlier version.

The results of HCC image classifications are summarized in Table 3.3–Table 3.5, having biopsy samples, surgical samples and a mix of biopsy and surgical samples as the training data respectively. The tables show that the new features did not give better results when the features were utilized separately. However, combinations of the new features with the nuclear and structural features could improve the accuracy for about 1–3% depending on the type of training and test data. For example, in biopsy samples, the sensitivity is improved from 84.7% to 88.2% while the specificity is unchanged (91.9%).

In the meantime, classifying biopsy samples by using surgical samples as the training data and vice versa gave lower results compared to if the samples were classified by using the same sample type. In the meantime, the classification rates increased when both biopsy and surgical samples were employed as training data. Nonetheless, the improvement on the classification rates might also be influenced by the increment of the

number of training data, since data from both biopsy and surgical resected samples were used as training data.

The tables also show that masking the images gave effect on the classification rates. However, the effect was different between biopsy and surgical samples. The classification results were improved when classifying the biopsy samples while decreased in the case of surgical samples.

Table 3.3 Experiment results using various set of features with biopsy resected samples as the training data.

Test Data	Set of features	ID	Sensitivity	Specificity
Biopsy sample	78 nuclear (unmasked)	BB01	86.36%	88.29%
	78 nuclear (masked)	BB02	88.00%	91.47%
	72 nuclear (unmasked)	BB03	85.09%	87.90%
	72 nuclear (masked)	BB04	86.91%	90.28%
	11 tissue changes	BB05	67.82%	72.82%
	10 trabecular	BB06	76.00%	84.52%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	BB07	88.55%	90.87%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	BB08	88.18%	91.67%
Surgery sample	78 nuclear (unmasked)	BS01	100.00%	10.88%
	78 nuclear (masked)	BS02	99.45%	5.44%
	72 nuclear (unmasked)	BS03	98.53%	29.08%
	72 nuclear (masked)	BS04	96.50%	30.21%
	11 tissue changes	BS05	48.43%	77.30%
	10 trabecular	BS06	88.40%	64.17%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	BS07	99.45%	39.40%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	BS08	94.11%	45.40%

Table 3.4 Experiment results using various set of features with surgical resected samples as the training data.

Test Data	Set of features	ID	Sensitivity	Specificity
Biopsy sample	78 nuclear (unmasked)	SB01	76.18%	91.27%
	78 nuclear (masked)	SB02	76.73%	86.11%
	72 nuclear (unmasked)	SB03	65.45%	90.87%
	72 nuclear (masked)	SB04	74.55%	84.72%
	11 tissue changes	SB05	60.36%	46.83%
	10 trabecular	SB06	62.18%	86.71%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	SB07	45.09%	99.01%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	SB08	76.18%	93.85%
Surgery sample	78 nuclear (unmasked)	SS01	88.21%	87.99%
	78 nuclear (masked)	SS02	87.48%	87.24%
	72 nuclear (unmasked)	SS03	90.61%	88.37%
	72 nuclear (masked)	SS04	87.66%	87.05%
	11 tissue changes	SS05	79.19%	73.36%
	10 trabecular	SS06	76.61%	86.68%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	SS07	93.19%	90.62%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	SS08	90.98%	88.56%

Table 3.5 Experiment results using various set of features with a mix of surgical and biopsy resected samples as the training data.

Test Data	Set of features	ID	Sensitivity	Specificity
Biopsy sample	78 nuclear (unmasked)	CB01	84.73%	91.87%
	78 nuclear (masked)	CB02	85.27%	90.67%
	72 nuclear (unmasked)	CB03	83.45%	89.68%
	72 nuclear (masked)	CB04	84.55%	89.68%
	11 tissue changes	CB05	67.45%	60.71%
	10 trabecular	CB06	74.36%	86.71%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	CB07	88.18%	91.87%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	CB08	88.18%	91.87%
Surgery sample	78 nuclear (unmasked)	CS01	88.95%	87.62%
	78 nuclear (masked)	CS02	89.50%	87.62%
	72 nuclear (unmasked)	CS03	91.16%	87.99%
	72 nuclear (masked)	CS04	90.06%	87.05%
	11 tissue changes	CS05	69.06%	75.05%
	10 trabecular	CS06	82.50%	80.86%
	72 nuclear (unmasked) + 10 trabecular + 11 tissue changes	CS07	93.37%	89.87%
	72 nuclear (masked) + 10 trabecular + 11 tissue changes	CS08	91.34%	89.68%

Figure 3.8 shows the receiver operating characteristics (ROC) curves when biopsy samples having a different set of features are used as the test data along with their sensitivity values when specificity is 90%. The values of area under curves (AUC) are as follow.

- AUC = 0.941 in the case of 78 unmasked nuclear with biopsy samples as the training data (BB01).

- AUC = 0.940 in the case of 78 unmasked nuclear with a mix of biopsy and surgical samples as the training data (CB01).
- AUC = 0.940 in the case of 78 masked nuclear with a mix of biopsy and surgical samples as the training data (CB02).
- AUC = 0.960 in the case of a combination of 72 masked nuclear and 21 new features with a mix of biopsy and surgical samples as the training data (CB08).

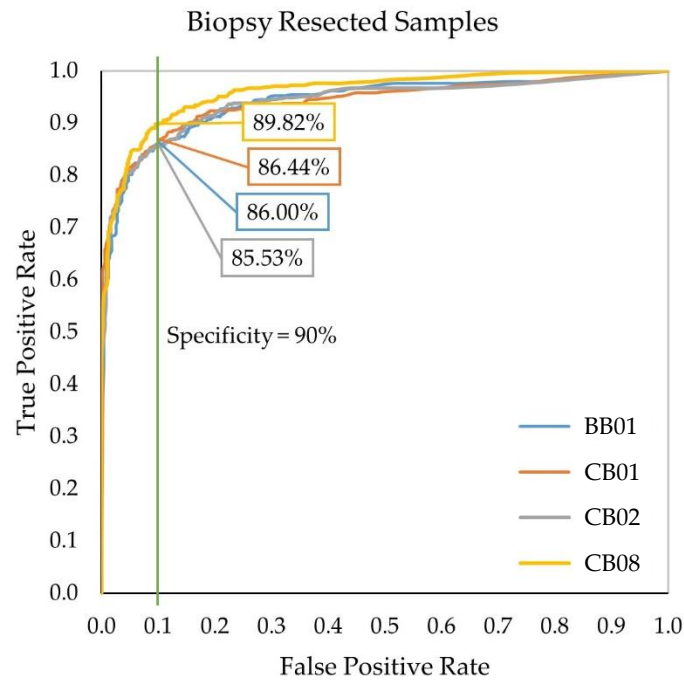


Figure 3.8 ROC curves for HCC classification on experiment BB01, CB01, CB02, and CB08 [17].

On the other hand, Figure 3.9 shows the ROC curves in the case of surgical samples, having AUC values as follow.

- AUC = 0.957 in the case of 78 unmasked nuclear with surgical samples as the training data (SS01), which was the similar condition with the method by Kiyuna et al.
- AUC = 0.955 in the case of 78 unmasked nuclear with a mix of biopsy and surgical samples as the training data (CS01).
- AUC = 0.951 in the case of 78 masked nuclear with a mix of biopsy and surgical samples as the training data (CS02).

- AUC = 0.968 in the case of a combination of 72 masked nuclear and 21 new features with a mix of biopsy and surgical samples as the training data (CS08).

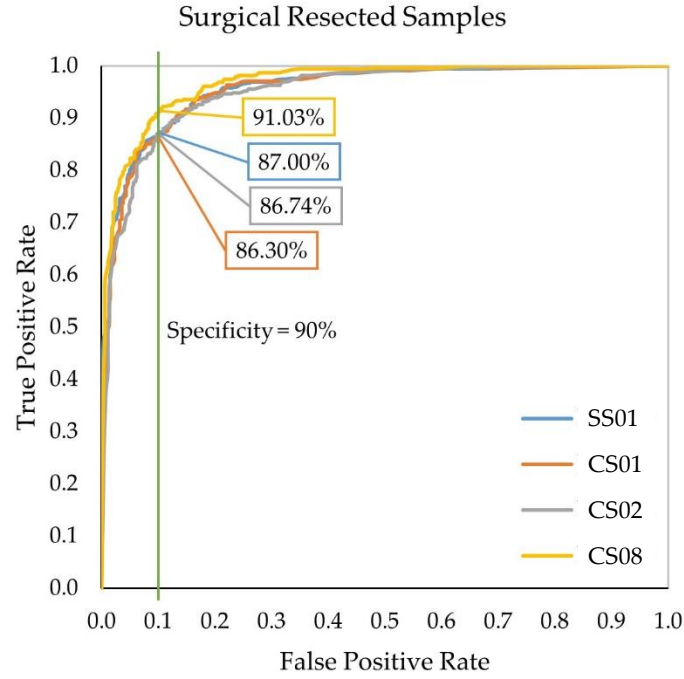


Figure 3.9 ROC curves for HCC classification on experiment SS01, CS01, CS02, and CS08 [17].

Further analysis was performed by comparing the result of the proposed method with the result from the method proposed by Kiyuna et al. in [14]. The condition on each method are shown in Table 3.6. The focus of the analysis was to evaluate the effectiveness of the proposed method on HCC detection for each HCC grade. The analysis was performed by comparing the true positive rates on each HCC grade between the two methods from the surgical samples, as shown in Table 3.7. Here, ROI images having 2 or more grades detected in an image were removed for the comparison and only use images that are labeled to one of the grades in EGS. From the table, the detection rates for the Grades I, II, and III are increased. The well-differentiated type of tumor shows the highest improvement that is, from 65.0% to 77.5%, by the proposed method. This improvement is important since the tissue structure in well-differentiated tumor is the most similar to the normal liver tissue compared to the other grades, which means HCC images from Grade I are more likely to be misclassified as non-cancer. The reduction of

detection rate in the case of undifferentiated tumor is likely due to the small amount of data available in this grade.

Table 3.6 Details of condition between the method by Kiyuna et al. [14] and the proposed method, in the case of surgical resected samples as training data.

No.	Condition	Method by Kiyuna	Proposed Method
1.	Training data	Surgical	Surgical
2.	Removal of non-hepatocyte nuclei through masking process	No	Yes
3.	Features used:		
	- 11 nuclear features (with 6 statistical distribution for each feature)	Yes	Yes
	- nuclei density of structural features (with 6 statistical distribution)	Yes	Yes
	- trabecular thickness of structural features (with 6 statistical distribution)	Yes	No
	- 11 tissue changes features	No	Yes
	- 10 trabecular features	No	Yes

Table 3.7 Comparison on true positive rates between Kiyuna et al. [14] method and the proposed method, in the case of surgical resected samples as training data.

Grade	Previous Method	Proposed Method
Grade 1.0	65.00% (52/80)	70.00% (56/80)
Grade 2.0	91.29% (262/287)	93.38% (268/287)
Grade 3.0	96.88% (62/64)	100% (64/64)
Grade 4.0	95.45% (21/22)	100% (22/22)
Overall	87.64% (397/453)	90.51% (410/453)

3.5. HCC image classification based on tissue types

In this section, another approach is introduced to improve the classification rates of HCC images. HCC will lead to architectural pattern and cytological changes. Some structural and tissue changes features used for the HCC classification are not suitable to be utilized on certain tissue type due to its characteristics. Here, a method to classify HCC images based on the tissue types is introduced. The result of tissue classification will then be used on the HCC classification where the features used in the process will depend on the tissue type of the HCC image.

3.5.1. Types of HCC tissue histological pattern

In the case of HCC, there are changes in the tissue morphologically. The changes may be found both in the pattern's architectures and in the cell form (cytologically), both may occur in combination. For instance, the nuclei will grow larger and become more heterogeneous. The number of nuclei in a unit area is increased as the cells are multiplied. The cells also often contain fat. Moreover, while hepatocytes are arranged in one line on the healthy liver tissue, the trabecular structure will become more irregular, and the cells are multiplied to 2-3 lines. WHO has described several HCC histologic patterns and cytological variants.

Trabecular (plate-like), pseudoglandular (acinar or adenoid), and compact (solid or sheet) pattern are the typical patterns founded in HCC cases. Trabecular pattern, or also called as the sinusoidal pattern is the most common pattern in liver parenchyma, especially in early-stage tumors. Here, the trabecular grow and become thicker, having twice or more lines compared to the normal tissue, and separated by sinusoid-like blood spaces. Pseudoglandular is less frequent compared to the trabecular pattern, showing the gland-like pattern and often found together with the trabecular pattern. Sheet pattern is relatively uncommon. Here, the tissue is formed by sheets of tumor cells and difficult to see the trabecular pattern. Other than the three mentioned pattern, some particular pattern can also be found such as the scirrhous pattern. Here, fibrous connective tissue becomes prominent, separating the tumor cell nests, and it may be associated with the other HCC patterns such as trabecular, pseudoglandular or sheet. In the case of cytological variants, there are pleomorphic cells, clear cells, spindle cells, fatty change, bile production, hyaline bodies, pale bodies and ground glass inclusions. Figure 3.10 below shows some architectural patterns and cytological variants of HCC.

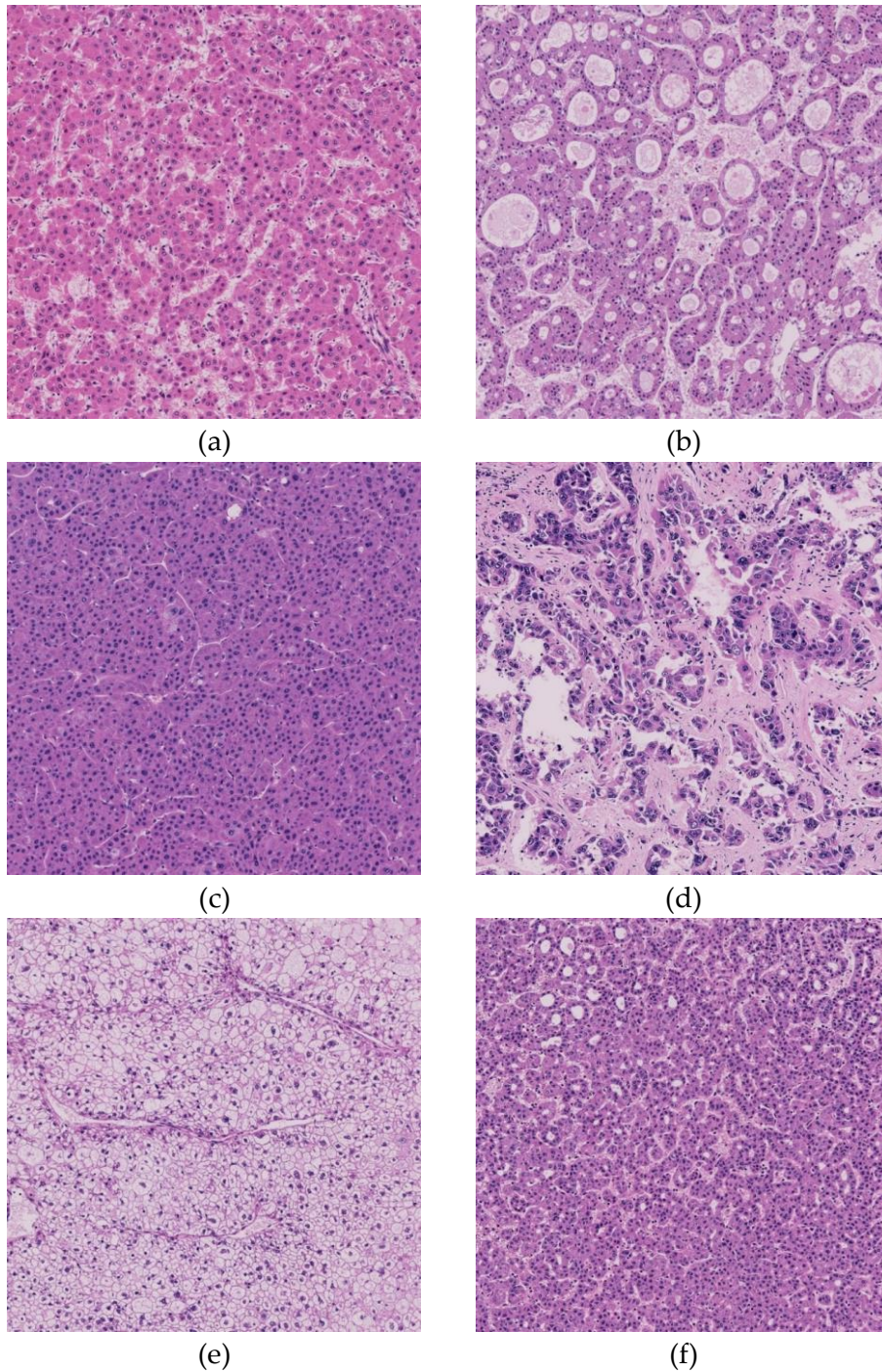


Figure 3.10 Various types of HCC tissue histologic pattern; (a) Trabecular (or sinusoidal) pattern, (b) Pseudoglandular (or acinar, adenoid) pattern, (c) Sheet (or solid, compact) pattern, (d) Scirrhous pattern, (e) Clear cell. The cytoplasm becomes clear due to fat, glycogen or water, (f) Trabecular pattern containing area with pseudoglandular pattern.

3.5.2. Overview of classification system based on tissue type

The overview system of HCC image classification based on the tissue type is shown in Figure 3.11. In the process of HCC image classification earlier, some set of features are extracted from ROI images that have been masked by the six types of the mask image. The features are then used to classify the ROI images either cancer or non-cancer through HCC classification process. Here, we extend the system by embedding the tissue classification process (marked as light blue blocks). In this work, we are focusing on classifying the tissue into two types only, either all structural and tissue changes features can be applied to the tissue type for HCC classification or not. Images that are suitable to use the structural and tissue changes features for the HCC classification are labeled as *trabecular type* while images that are only suitable to use nuclear features for the HCC classification are labeled as *non-trabecular type*.

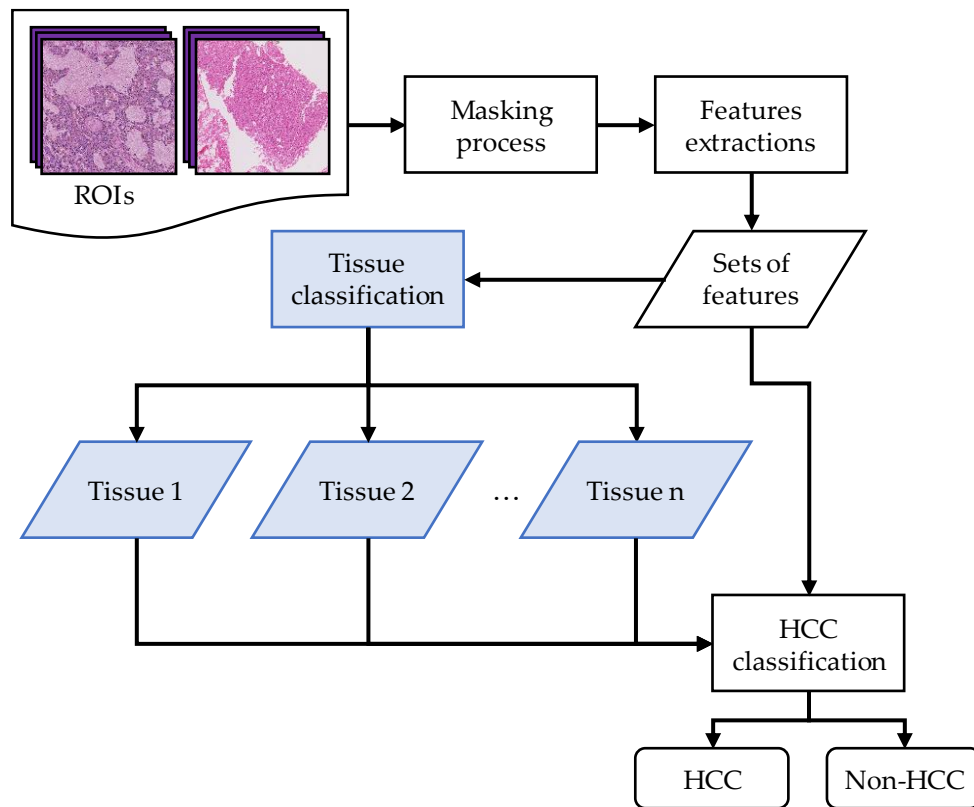


Figure 3.11 Diagram of ROI images classification based on the tissue type.

3.5.3. Texture-based tissue features of liver

To classify the tissue type, a statistical method to examine the image texture known as grey level co-occurrence matrix (GLCM) is employed. GLCM accounts the spatial relationship between pixels, calculating the frequency of pairs of pixels having a specific value on certain spatial relationship occurred in an image. Based on this relationship, four statistical parameters are derived.

- 1) *Contrast*, measuring the intensity contrast between a pixel and its neighbor over the whole image. Mathematically, it is defined as:

$$\sum_{i,j} |i - j|^2 p(i, j)$$

- 2) *Correlation*, measuring the correlation of a pixel to its neighbor over the whole image. Mathematically, it is defined as:

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$$

- 3) *Angular second moment*, calculating the sum of squared elements in the GLCM. It is defined as:

$$\sum_{i,j} p(i, j)^2$$

- 4) *Homogeneity*, measuring the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It is defined as:

$$\sum_{i,j} \frac{p(i, j)}{1 + |i - j|}$$

In defining the spatial relationship, also known as the *offset*, four neighboring pixels having different angles (0°, 45°, 90°, and 135°) to the pixels of interest are used as illustrated in Figure 3.12.

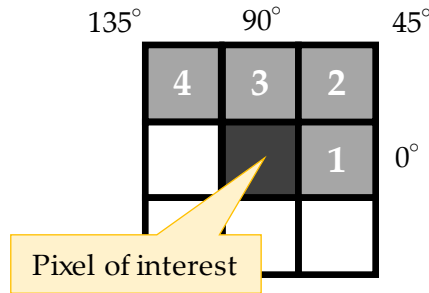


Figure 3.12 Four offsets used in extracting the GLCM properties.

The texture features are extracted from ROI images that have been modified by masking the images with nuclei mask, fiber mask, glass mask, RBC mask, and sinusoid mask. The black color is used in the masking the nuclei while white color is used in the other masks. The fat mask is not included here since the fat drop is considered as part of the cytoplasm in the liver cells. As GLCM works in grey-level spatial, the features can only be extracted from the grayscale image of the ROI image. For efficiency, the green channel of the ROI image is utilized since the peak absorptions of both hematoxylin and eosin stain are in the green wavelength as shown in Figure 3.13. To increase the number of data and to prevent processing on the large image, each masked image is divided into 4x4 areas, resulting in 16 sub-ROIs (512x512 pixels) for each image. Furthermore, for each sub-ROI image, another three images having different resolutions (256, 128 and 64 pixels) were generated to get more understanding on the texture of the mask area. In resizing the sub-ROI image, a bicubic interpolation method is utilized. Here, 64 texture features have been extracted (4 GLCM parameters $\times$ 4 offsets $\times$ 4 resolutions). The average and standard deviation from the four offsets are calculated to reduce the number of features, resulting 32 extracted texture features. Figure 3.14 illustrated the process in generating the green channel of the ROI images while Figure 3.15 shows edge shape of masking in the four resolution images.

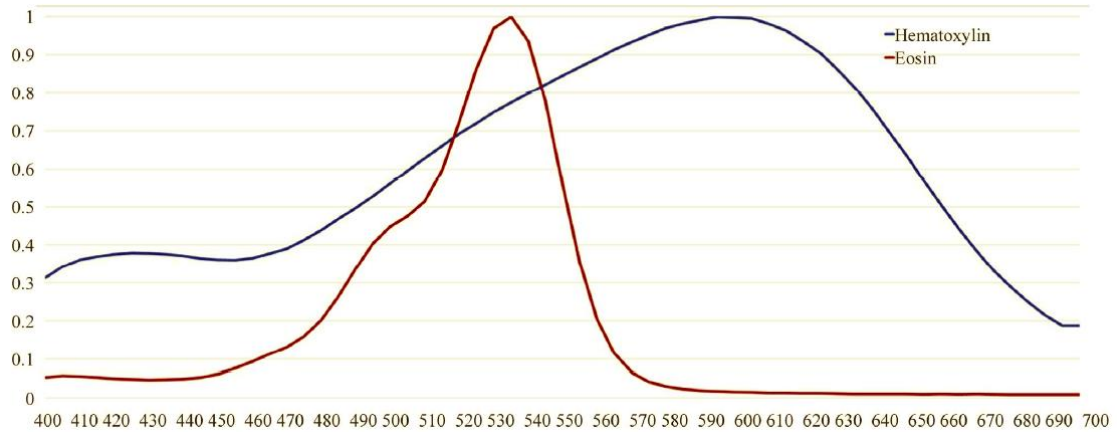


Figure 3.13 Absorption of hematoxylin and eosin stain in samples. The peak absorptions of hematoxylin and eosin are at 590 nm and 530 nm respectively. These are around the wavelength of green color (546.1 nm) compared to blue (435.8 nm) and red (700 nm) [34], [54].

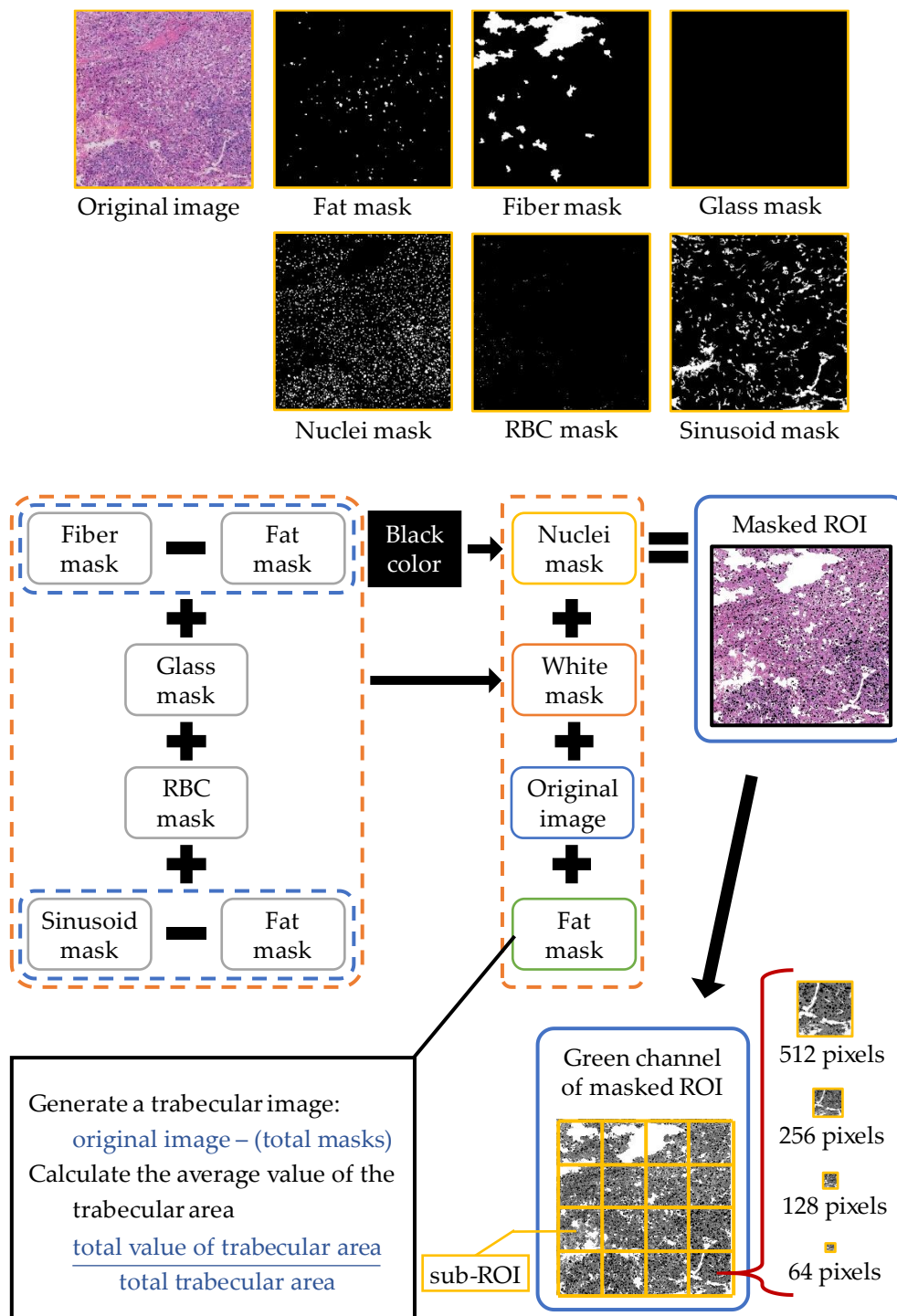


Figure 3.14 Process in generating image for texture features extraction.

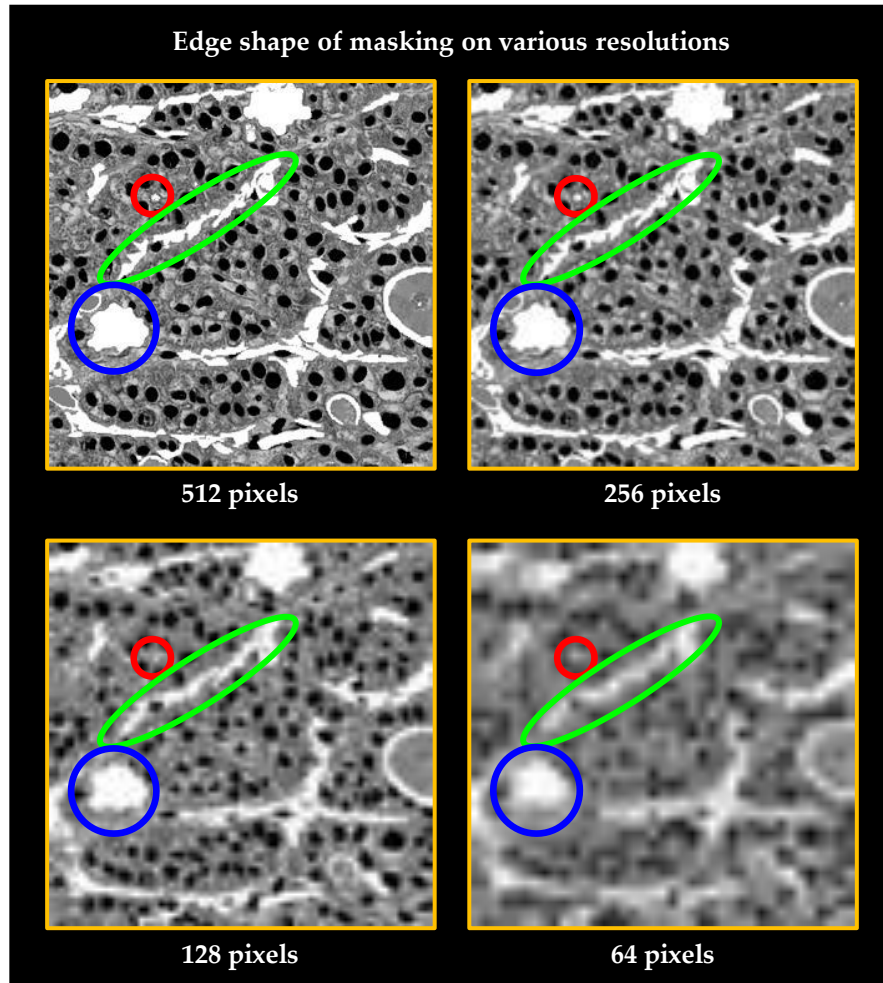


Figure 3.15 Sub-ROI with different resolutions provides information on the image texture. In the above images, three different regions are highlighted; small area (red), long but thin area (green) and large area (blue). The small area is “gone” in low resolution image while large area can still be seen although with blur effect in the edge.

3.5.4. Liver image classification based on tissue type

The classification process used 100 ROI images of surgery resected samples as training data, where 47 images were the trabecular type. As for the test data, 140 ROI images were selected. Seventy ROI images have been commented by the experienced pathologists, where 46 were trabecular or related to trabecular. The other 70 images for the test data were selected from images having non-cancer cases.

Before the classification process, the value of each feature was scaled in the range of [0,1]. Moreover, images covered with too much mask were filtered to improve the classification reliability. Here, we chose 60% white mask on the image with 256 pixels as the threshold. If the white mask on the image is more than 60%, the image will be removed from the classification process (Figure 3.16). In the end, 92 sub-ROI images were removed from the training data and 95 sub-ROI images were removed from the test data, resulting 1508 and 2145 sub-ROI images for training and test data respectively.

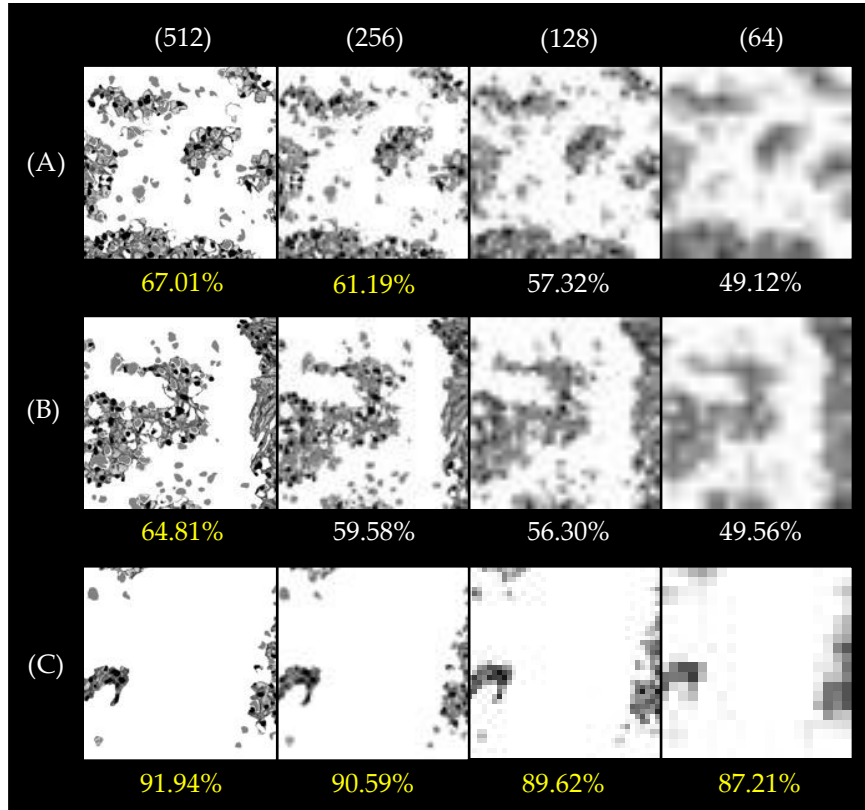


Figure 3.16 Filtering sub-ROIs. Images having too much white area are removed from the data. Image (a), (b), and (c) are covered with 61.90%, 59.58% and 90.59% of white mask respectively. Image (a) and (c) were removed from the classification process.

For the classification process, support vector machine (SVM) was utilized. Radial basis function was used as the kernel type, and LibSVM was utilized as the library. Five-fold cross-validation was used to evaluate the classification performance. In Table 3.8, the classification results on the sub-ROI classification are shown. Meanwhile, Table 3.9 shows the classification results based on the ROI images. Here, the class of an ROI image is determined based on the following rule: if the number of sub-ROI in an ROI

predicted as trabecular is more than or equal to non-trabecular, then the ROI will be classified as trabecular. Else, the ROI will be classified as non-trabecular. Figure 3.17 illustrates the classification of the ROI images. Sub-ROI regions overlaid by blue color are classified as trabecular type while sub-ROI regions overlaid by red color are classified as non-trabecular type. In the case of Figure 3.17(a), only 7 of the sub-ROI regions were classified as trabecular; thus the ROI was classified as non-trabecular type. While in Figure 3.17(b), the number of sub-ROI regions classified as the trabecular type was equal with the non-trabecular type; thus the ROI was classified as trabecular type.

Table 3.8 Classification results on sub-ROI images based on the pattern types.

		Real Condition	
		Trabecular	Non-trabecular
Classification Result	Trabecular	798	488
	Non-trabecular	452	407
	Total	1250	895
Accuracy		56.18%	

Table 3.9 Classification results on ROI images based on the pattern types.

		Real Condition	
		Trabecular	Non-trabecular
Classification Result	Trabecular	58	36
	Non-trabecular	23	23
	Total	81	59
Accuracy		57.86%	

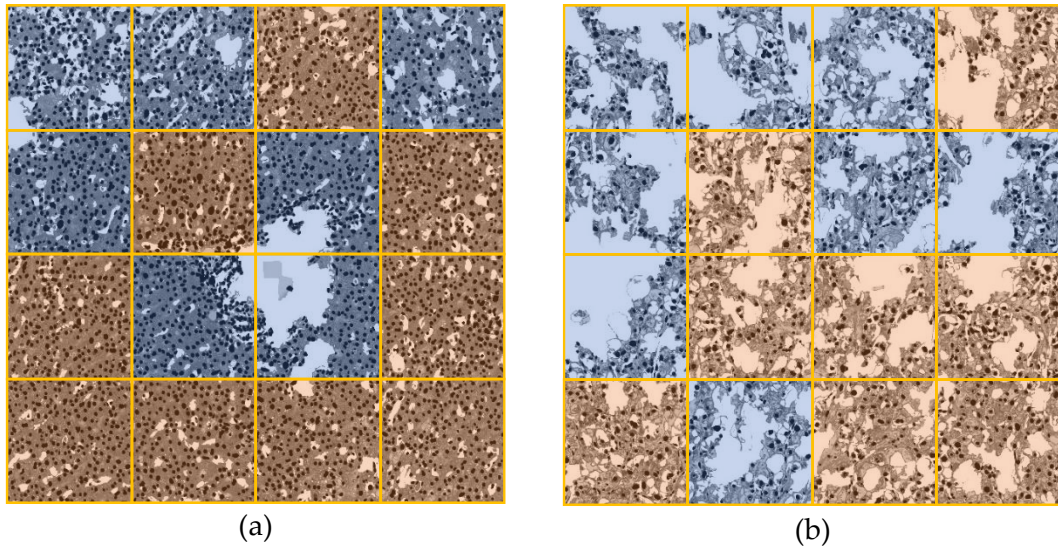


Figure 3.17 Classification results on two ROI images.

Here, the accuracy of tissue classification based on texture features was not high. Nevertheless, by analyzing the images, the structural features might be useful in the process of cancer classification on some non-trabecular images that were predicted as trabecular tissue and vice versa. Figure 3.18 shows a cord-like tissue (non-trabecular type) that was classified as trabecular by the system. By studying the masked image, we believe that the structural features are useful in HCC image classification in this case.

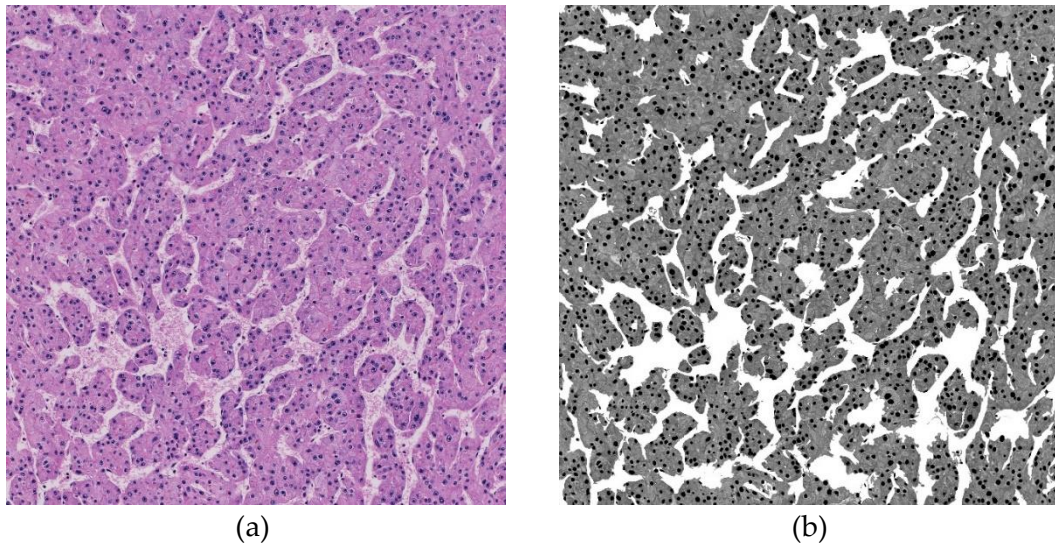


Figure 3.18 ROI classified as cord-like tissue type by pathologist is classified as trabecular tissue type by the tissue classification.

3.5.5. HCC image classification post tissue classification

The result of image classification based on tissue type was used for HCC classification on HCC images as shown in Figure 3.19. Images that had been classified as trabecular used all features (72 nuclear, 11 trabecular and 10 tissue changes features) while images that had been classified as non-trabecular used nuclear features only. For the training data, 615 ROIs of surgery samples were selected. Again, SVM with radial basis function kernel was employed. LibSVM was utilized as the library, and five-fold cross-validation was used to evaluate the performance.

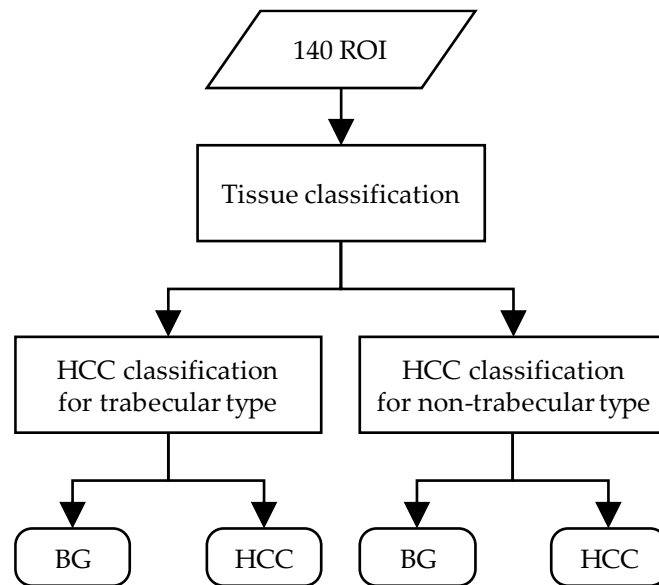


Figure 3.19 Diagram of HCC classification based on the tissue types.

Figure 3.20 shows the result of HCC classification in this experiment. Cases A and D are the proposed method. In case A, ROI images that were classified as the non-trabecular type used the nuclear features only for the HCC classification. In case of D, ROI images that were classified as the trabecular type used the nuclear and structural features for the HCC classification. Cases B and C are shown to compare the results. From the table in the figure, case A (97.78% accuracy) were more accurate compared to case B (93.33% accuracy), showing that HCC classification of non-trabecular type tissues are more suitable to use the nuclear features only. On the other hand, HCC classification on trabecular type tissue gave the better result when nuclear and structural features are

utilized, as case D (86.96% accuracy) gave the better result than case C (85.87% accuracy). Generally, the results show that splitting HCC image classification based on the tissue type has potential to improve the classification rates.

		Non-trabecular			Trabecular			
Nuclei	Case A				Case C			
			Condition				Condition	
			HCC	Non-HCC			HCC	Non-HCC
	Class. Result	HCC	36	1	Class. Result	HCC	31	10
		Non-HCC	0	8		Non-HCC	3	48
Total		36	9	Total		34	58	
Accuracy		97.78%		Accuracy		85.87%		
Nuclei + structural	Case B				Case D			
			Condition				Condition	
			HCC	Non-HCC			HCC	Non-HCC
	Class. Result	HCC	36	3	Class. Result	HCC	32	10
		Non-HCC	0	6		Non-HCC	2	48
Total		36	9	Total		34	58	
Accuracy		93.33%		Accuracy		86.96%		

Figure 3.20 The result of HCC classification on ROI images that have been categorized between trabecular and non-trabecular type. Cases A and D are the proposed method, while case B and C are shown for comparing the result.

3.6. Discussion

Overall, the experiments show that the biopsy and surgical samples can be mixed and used together as training data in HCC image classification, although both samples have different characteristics. In the implementation, biopsy samples are more often to be utilized as the object of diagnosis since the sample is easier to be removed from the patients. However, relying on biopsy sample only for training data will cause limited number of data. In the meantime, surgically resected samples can provide more training data due to its large size. Moreover, features extracted from the surgical sample are more stable compared to biopsy samples since biopsy samples are often torn. Also, the variations of tissue type are shown clearly in surgical samples. Hence, the possibility to use both sample types together become essential in HCC image classification.

In the meantime, the effect of image masking was different between biopsy sample and surgical samples as some attributes are missed due to the masking process in surgical samples. Nonetheless, the masking process on the nuclear features facilitates the reliability of the nuclear features since falsely detected nuclei, such as lymphocytes or non-hepatocyte nuclei, are removed from the quantification. The masking process also brought effects on the distribution difference between cancer and non-cancer as shown in Figure 3.21. In case (a), the distribution difference between cancer and non-cancer was increased in 50 percentile of nuclei density of biopsy samples. Meanwhile, in case (b), the value range of masked images was reduced in 50 percentile of nuclei density of surgical samples. More details on the distribution differences on each nuclear feature before and after the image masking are shown in Appendix A.

The experiments also demonstrated that the combination of nuclear, trabecular, and tissue changes features would increase the classification rates in HCC detection using SVM. In fact, the detection of lower grade tumor was improved by the combination of features. This is because the nuclei shapes in the case of well-differentiated tumors are similar to normal liver, yet there are deformations in the arrangement of cell pattern. Therefore, appending the new features to the classification process can expand the difference between the normal tissues and the tissues having low-grade cancer. Even so, the rate of undifferentiated tumors is worse. As the state of cancer is getting worst, the sinusoid areas become disappear and thus the sinusoid segmentations for undifferentiated tumors are sometimes inaccurate. Moreover, the lack of samples in this grade may also contribute to the misclassification. Further works are required to

investigate the grade using more training data and to improve the sinusoid segmentation. Overall, the proposed methods improve the chance of early HCC detection, which is very important and has been one of the main challenges in the area, even though it should be noted that the improvement on the classification rates in this study might also be affected by the increment number of training data since both biopsy and surgical resected samples were used as the training data.

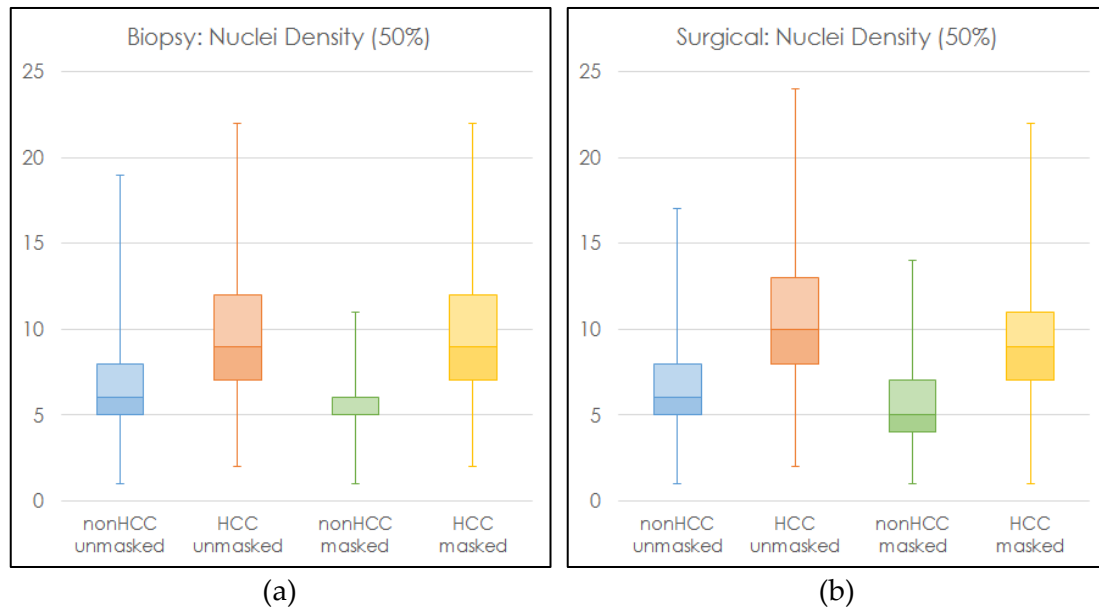


Figure 3.21 Effect of masking on distribution difference of nuclear feature.

On the other hand, image classification based on the tissue type ahead of HCC image classification has potential to improve the classification rates. The results on the experiment also indicate that the features related to tissue structure are dependent to the tissue type, while nuclear features are more robust. More data on this experiment may deliver more understanding on the effect of tissue classification based in the tissue type on HCC image classification.

Chapter 4

Color correction and its effect on quantitative pathology

This chapter focuses on the evaluation on the effectiveness of color correction techniques on WSIs acquired by different devices. First, a brief explanation on color correction methods are introduced. Then, an investigation on the effectiveness of a stain-based color correction method that earlier had shown its effectiveness when correcting WSI obtained by a single WSS is reported. It follows by another investigation the effectiveness of device-based color correction and a combination of device- and stain-based color correction on WSIs obtained by a set of WSSs.

4.1. Color correction on histopathological images

The primary purpose of color correction is to shift the color distribution of an image to the color distribution supposed to be by an ideal staining and ideal detector. Generally, there are two approaches for color correction on histopathological images which are device-based and stain-based corrections [55].

4.1.1. Device-based color correction

Color calibration is a common technique focusing on color difference on image captured by different devices. In principal, device-based color correction approach often uses a color chart slide which contains several color patches for the characterization of WSS's. In this approach, the color correction is performed by using color correction matrix derived from the target and scanned colors of the color patches. Various techniques of color calibration have been established in various fields, including in printing industry, digital camera, and radiology. However, in digital pathology itself, color calibration is still in its early development. Some works on the development of color calibration to correct color of WSIs captured by different scanning devices can be refer in [55]–[57]. Figure 4.1 (a) shows the workflow of the device-based color correction by Bautista et al. [55].

4.1.2. Stain-based color correction

Murakami et al. have introduced the stain-based color correction method in [51], in order to reduce the color difference on WSI due to staining process for further analysis of the WSI. This method has been implemented on a prototype system developed for liver biopsy samples focusing on hepatocellular carcinoma (HCC) detection. The stain-based color correction method intends to adjust the stain strength using Beer-Lambert law. In the stain-based technique, the color distribution of a WSI is analyzed, in which some regions that have representative color distribution are sampled. For each WSI, color distribution plane is generated in logarithm RGB color space by using the sampled color. Here, we suppose tissue sampled stained with hematoxylin and eosin (H and E), and three primary vectors (H, E, and residual) are identified based on the distribution. The color correction phase is then accomplished by representing the color of each pixel with the weighted sum of the primary vectors, in which two of the primary vectors (H and E) are replaced with the reference primary vectors and the staining strengths are adjusted to those of reference samples. The stain-based color correction method was shown to be effective in digitized specimen obtained with a single WSS. Figure 4.1 (b) shows the workflow of the stain-based color correction by Murakami et al. [51].

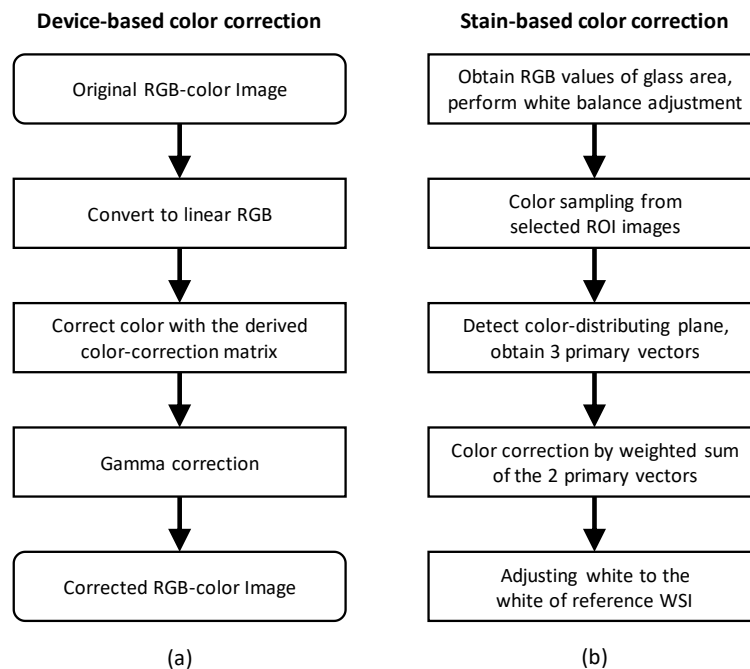


Figure 4.1 The workflows of the device- and stain-based color corrections introduced in [55] and [51].

4.2. Evaluation of stain-based color correction on histopathological images acquired by different WSSs

This section focuses on the investigation of the effect of stain-based color correction introduced by Murakami et al. [51] when the method was applied on WSIs captured by different devices. By comparing the quantified features of nuclei from the WSIs, both without and with color correction implementation, the effectiveness of the proposed stain-based color correction was evaluated. The nuclear features were chosen as the focus of this research as these features are well established and general as the nuclei extraction method had been applied to various type of tissue organs.

4.2.1. Experiment workflow

Three liver specimens were utilized in this work; a surgical specimen and two autopsy specimens, labeled as Sample 1, Sample 2, and Sample 3 as shown in Figure 4.2. The surgical specimen is a normal liver tissue while both autopsy specimens contain regions diagnosed as HCC. All samples were stained by H&E stain, where the cell nuclei become dark blue of hematoxylin and cytoplasm become pink as the eosin.

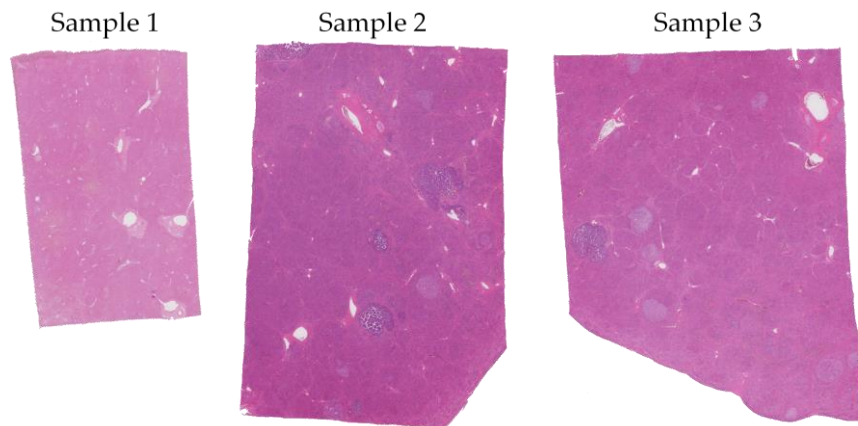


Figure 4.2 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs [58].

Whole slide images for the experiment were obtained by scanning the liver specimens using three WSSs; two slide scanners manufactured by Hamamatsu Photonics K.K. (NanoZoomer 2.0HT and NanoZoomer XR) and a slide scanner

manufactured by 3DHistech Ltd. (Pannoramic DESK). These scanners have their own output formats which depend on the manufacturer; NDPI for WSI captured by both NanoZoomer slide scanners and MRXS for 3DHistech. This difference can be overcome by implementing a WSI format converter. In addition to format difference, variations on sampling interval also existed on the produced WSIs regardless of the format. The sampling interval of WSI produced by NanoZoomer XR is 454 nm/pixel at 20 $\times$ magnification while it is 459 nm/pixel at the same magnification for NanoZoomer 2.0HT. In the meantime, 3DHistech Pannoramic DESK produces WSI having sampling interval 619 nm/pixel in x -direction and 620 nm/pixel in y -direction at 20 $\times$ magnification. In this work, these slide scanners are labeled as WSS-A, WSS-B, and WSS-C, while the produced images are labeled as WSI-A, WSI-B, and WSI-C respectively.

Two sets of 13 ROI images were selected from the WSIs at 20 $\times$ magnification and labeled as image A-M, where image A-D, image E-H, and image I-M were extracted from Sample 1, Sample 2, and Sample 3 respectively. A series of image transformation processes were then applied on the selected regions to equalize the sampling interval of the images, producing 1024 $\times$ 1024 pixels ROI images. The stain-based color correction technique described in [51] was then applied on one of the two ROI images sets, thus producing 78 images in total.

The uncorrected ROI images were compared with the color corrected version to confirm the effect of stain-based color correction on WSIs captured by different devices. To analyze the effect further, the CIELAB color values of 6 tissue components, which consist of nuclei, cytoplasm, lymphocyte, sinusoid, fiber, and red blood cell (RBC), were compared. In the process, a set of 3 $\times$ 3 image patches were extracted from each component. The mean RGB values on each patch were calculated, and the CIELAB color values were then derived, as described in [55]. The results were then projected onto the a - b chromaticity axes of the CIELAB color space.

To study the effect of color correction on the feature quantification, the prototype of histopathological image analysis software in [17] was utilized to extract morphological features from the ROI images. Note that the prototype system was developed based on WSI captured by WSS-A; thus ROI images from WSI-A were used as the reference. ROI images from captured by WSS-B and WSS-C needed to be converted and transformed so that they had the same format and sampling interval as images captured by WSS-A. Furthermore, since the prototype system was designed to process images of 2048 $\times$ 2048 pixels, the ROI images produced earlier were padded with white color so that its size

became 2048×2048 pixels. For each ROI images, a set of 93 tissue features were extracted, comprising nuclei, trabecular, and tissue changes features as described in [17]. The nuclear features consist of area, peripheral, circularity, ratio of short and long axis lengths of elliptic fit, cell contour complexity, and four gray-level co-occurrence matrix (GLCM) textures (angular second moment (ASM), contrast, homogeneity, and entropy). Each nuclear feature in an ROI image is represented by its six statistical distributions, which are 10th, 30th, 50th, 70th, 90th percentiles and standard deviation. The methods for feature calculation can be found in [14]. Note that the prototype system only detects nuclei from hepatocytes (liver cells), where the hepatocytes are detected by removing specific components (e.g., fiber and sinusoid), from the image through masking process as explained in [17]. Through the masking process, nuclei that do not belong to hepatocytes or other components that look like nuclei, such as lymphocytes, will not be included when quantifying the nuclear features. These whole processes are shown in Figure 4.3. The nuclear features from non-color corrected and color corrected images were then compared to investigate and analyze the effectiveness of the stain-based color correction method on feature quantification.

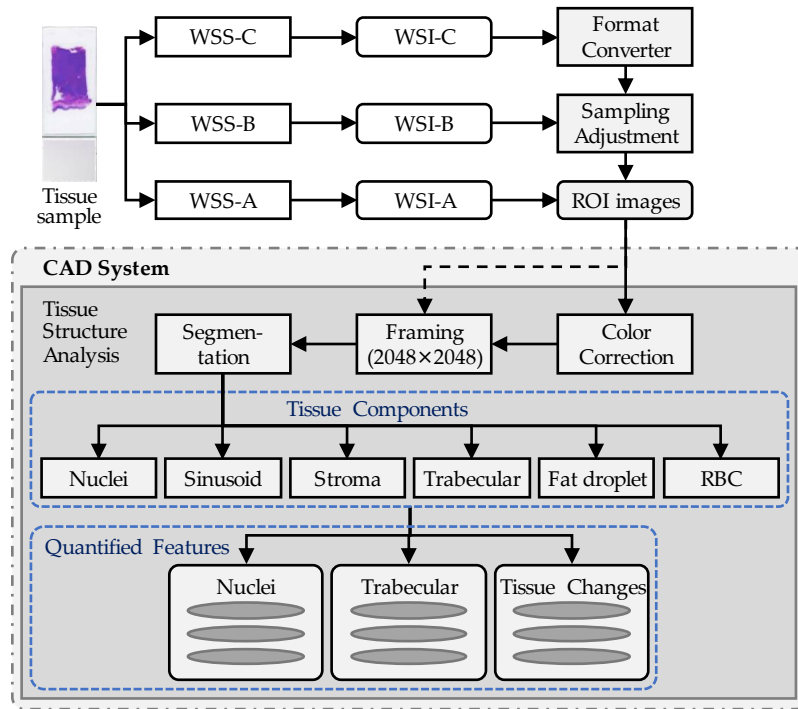


Figure 4.3 Workflow of experiment on stain-based color correction. 13 ROI images extracted from digitized of three tissue specimens used for the experiment, where each specimen was digitized thrice by three different WSSs [58].

4.2.2. Effect of stain-based color correction on tissue components

Figure 4.4 shows the effect of stain-based color correction on each ROI image where there is less color difference after implementing the color correction. Without color correction, the blue color on ROI images extracted from WSI-B is stronger compared to those from WSI-A and WSI-C, while the color on ROI images from WSI-C is brighter compared to the others. Figure 4.5 shows the projection of cluster plots of 6 tissue components on the a-b chromaticity axes of CIELAB color space. The WSI-A1, WSI-B1, and WSI-C1 labels refer to non-color corrected ROI images captured by WSS-A, WSS-B, and WSS-C respectively. On the other hand, WSI-A2, WSI-B2, and WSI-C2 labels are for the color-corrected version of the ROI images. Each tissue components were represented with 30 data samples, where the triangles represent the a and b color values of the tissue components before color correction while the circles represent the a and b color values of the tissue components after color correction. The circles are more compact and less scattered than the triangles, indicating that less color difference on the color of the tissue components.

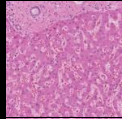
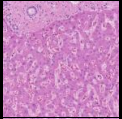
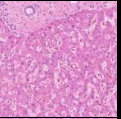
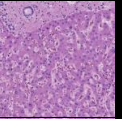
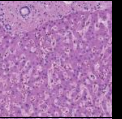
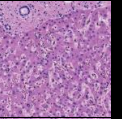
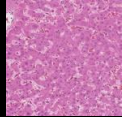
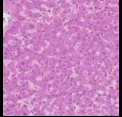
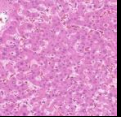
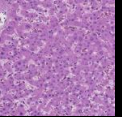
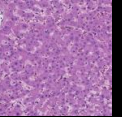
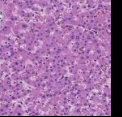
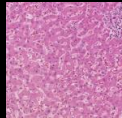
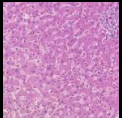
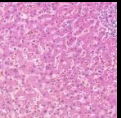
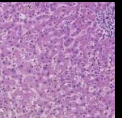
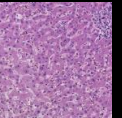
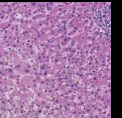
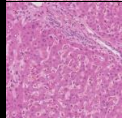
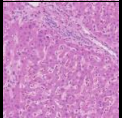
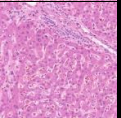
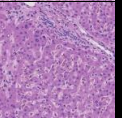
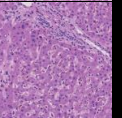
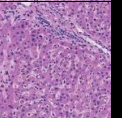
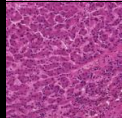
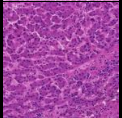
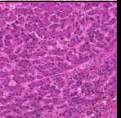
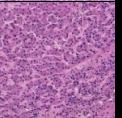
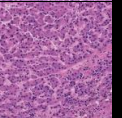
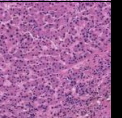
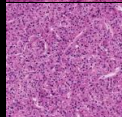
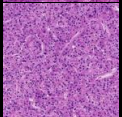
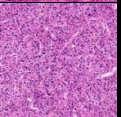
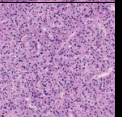
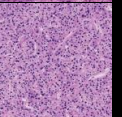
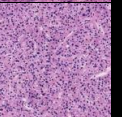
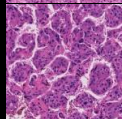
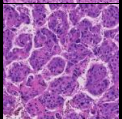
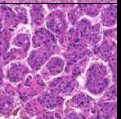
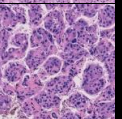
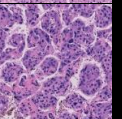
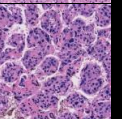
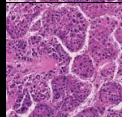
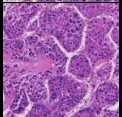
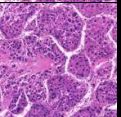
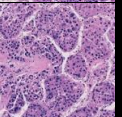
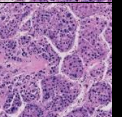
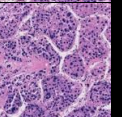
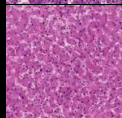
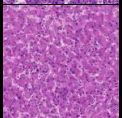
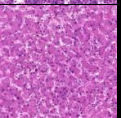
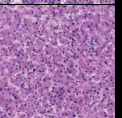
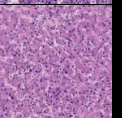
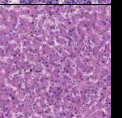
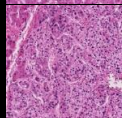
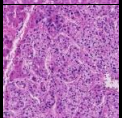
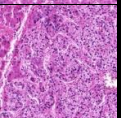
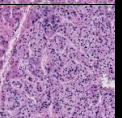
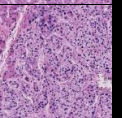
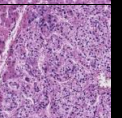
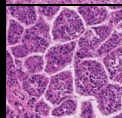
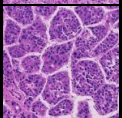
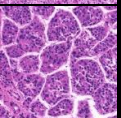
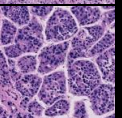
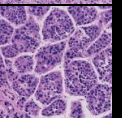
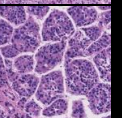
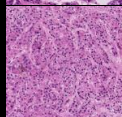
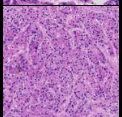
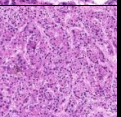
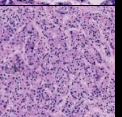
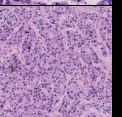
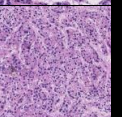
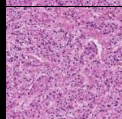
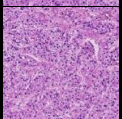
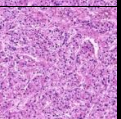
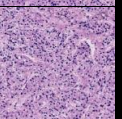
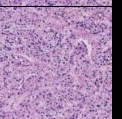
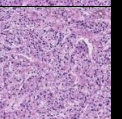
Image	w/o color correction			with color correction		
	WSS-A	WSS-B	WSS-C	WSS-A	WSS-B	WSS-C
A						
B						
C						
D						
E						
F						
G						
H						
I						
J						
K						
L						
M						

Figure 4.4 The effect of stain-based color correction on the three sets of 13 ROI image [58].

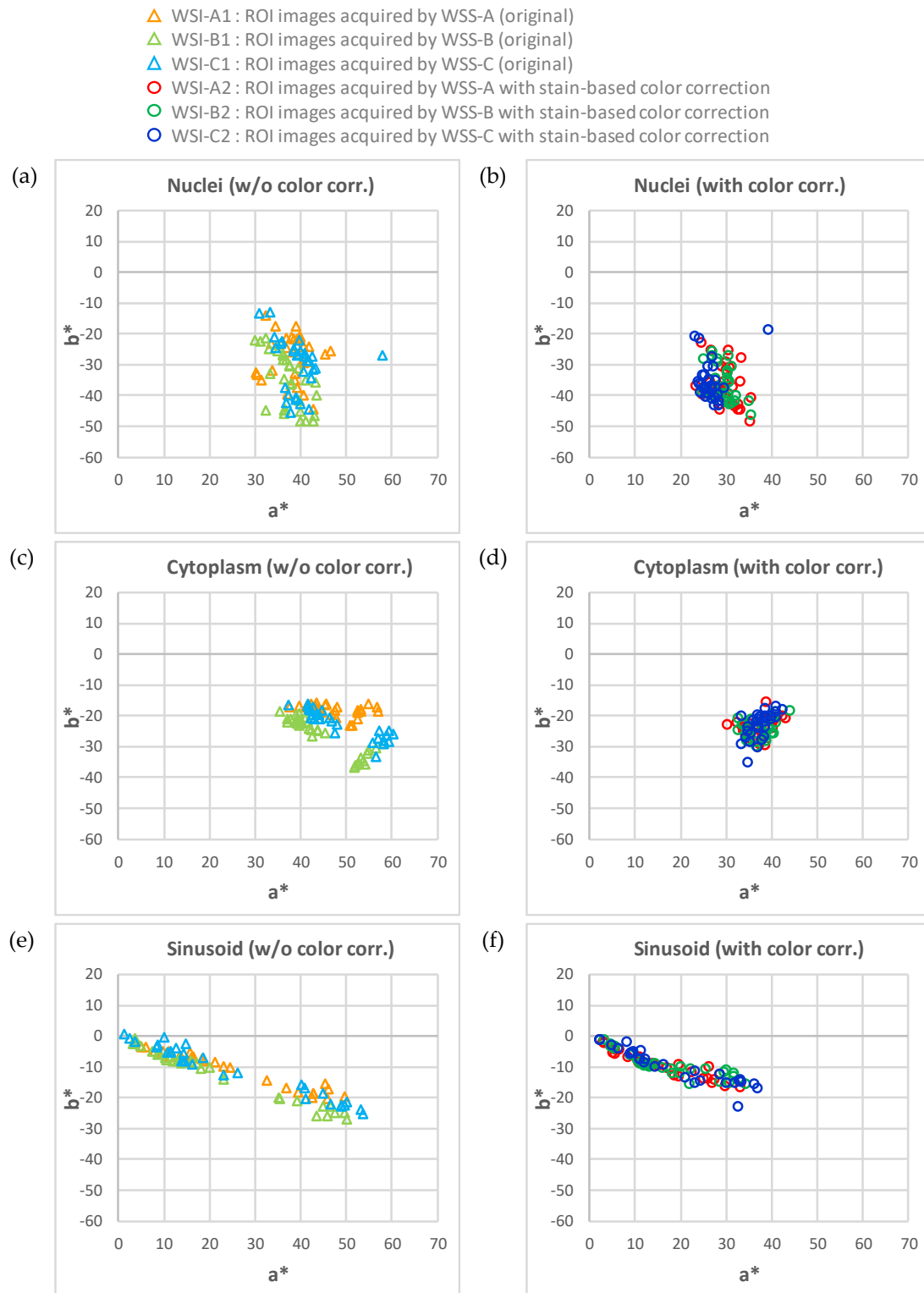


Figure 4.5 The projection of cluster plots of six tissue components (nuclei, cytoplasm, sinusoid, lymphocyte, fiber, red blood cell) on the a - b chromaticity axes of CIELAB color space [58].

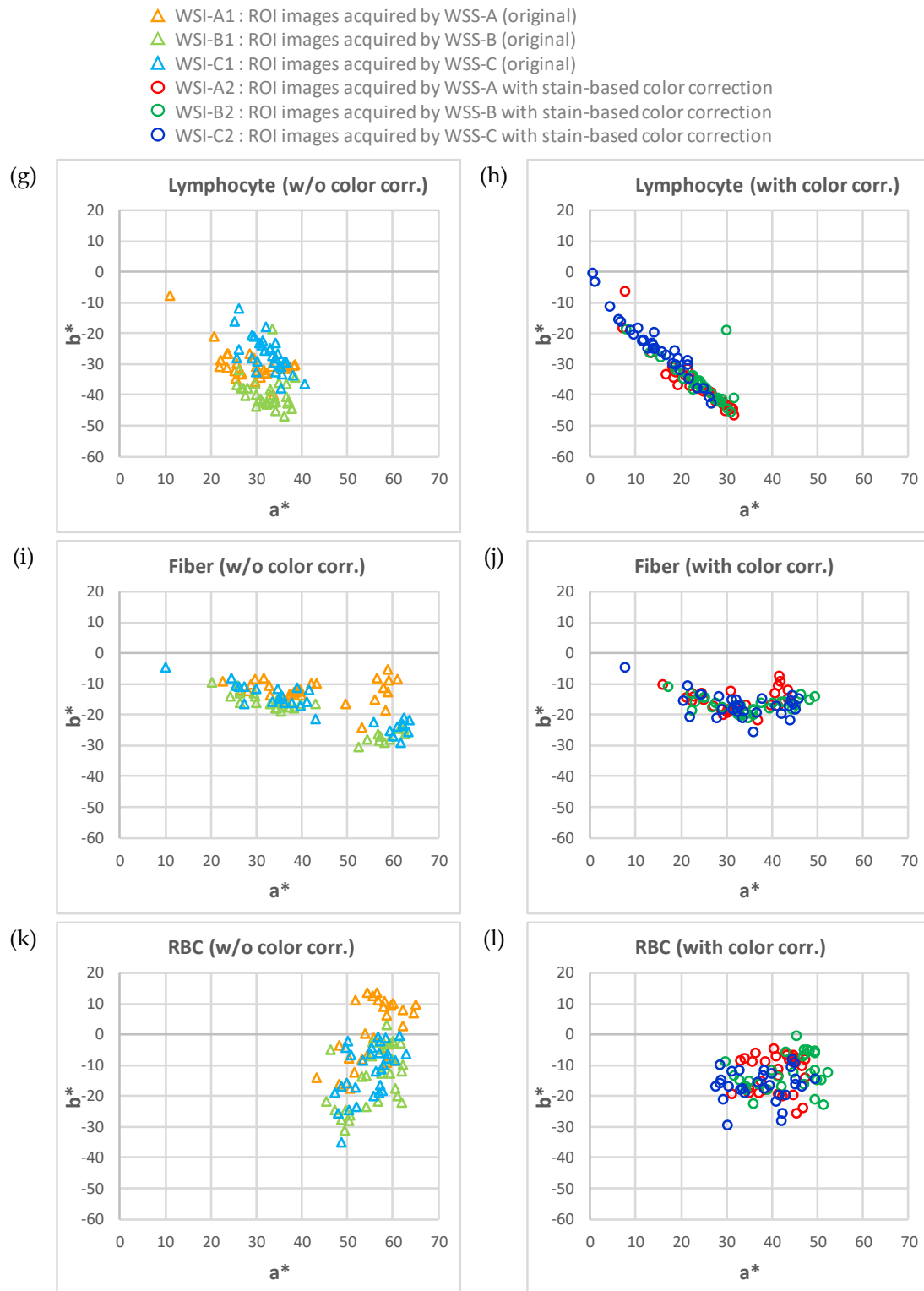


Figure 4.5. (continued).

4.2.3. Effect of stain-based color correction on nuclei detection

The effect of stain-based color correction on nuclei detection process is shown in Figure 4.6. As mentioned in Subsection 4.2.1, the stain-based color correction works for both stain variation and device dependence. Increment or reduction of the bar's height between upper and lower graphs shows the color correction effect on the stain variations. At the same time, the difference in the heights of three bars (the set of WSI-A, WSI-B, and WSI-C) shows the device dependency of the digitized specimen. The nuclei detections were failed and most of the nuclei were not detected before color correction in the case of ROI images extracted from Sample 1 (images A-D). The staining color on Sample 1 is not as strong as in Sample 2 and 3 in which the hematoxylin did not change the nuclei color blue enough to be detected as nuclei. In fact, most of the objects detected as nuclei on non-color corrected images of image A-D were actually lymphocytes, as the color of H&E stained lymphocytes is darker than nuclei. After the color correction, the numbers of detected nuclei are significantly increased, and this is the effect of correcting staining variation. Meanwhile, in the case of ROI images from Sample 2 and 3, the effect of stain-based color correction varies depending on the cases.

It is expected to compare the number of detected nuclei with true value, but in the prototype system of [18] the nuclear detection was not evaluated precisely including the comparison with true values. It is because only the statistical distribution was used as morphological features. Thus, we selected two images and the nuclear detection results were compared with the manual detections. Though there are some faint low contrast nuclei, only apparent ones are counted in the manual detection. Table 4.1 shows the results. After the color correction, the numbers of detected nuclei are larger than the manual count, because some non-hepatocyte nuclei and lymphocytes were detected as nuclei by the system and were not covered by the fiber or sinusoid masks. Figure 4.7 shows the example of manual and automatic detections. By looking at the images, the number of detected nuclei of hepatocytes is increased, along with the other objects such as lymphocytes. The lymphocytes and nuclei of other cells should be excluded, but the performance of the masking process previously explained is not sufficiently high. The improvement of the masking process is one of the future issues in the analysis system.

- WSI-A1 : Non-color corrected ROI images scanned by WSS-A
- WSI-B1 : Non-color corrected ROI images scanned by WSS-B
- WSI-C1 : Non-color corrected ROI images scanned by WSS-C
- WSI-A2 : Color corrected ROI images scanned by WSS-A
- WSI-B2 : Color corrected ROI images scanned by WSS-B
- WSI-C2 : Color corrected ROI images scanned by WSS-C

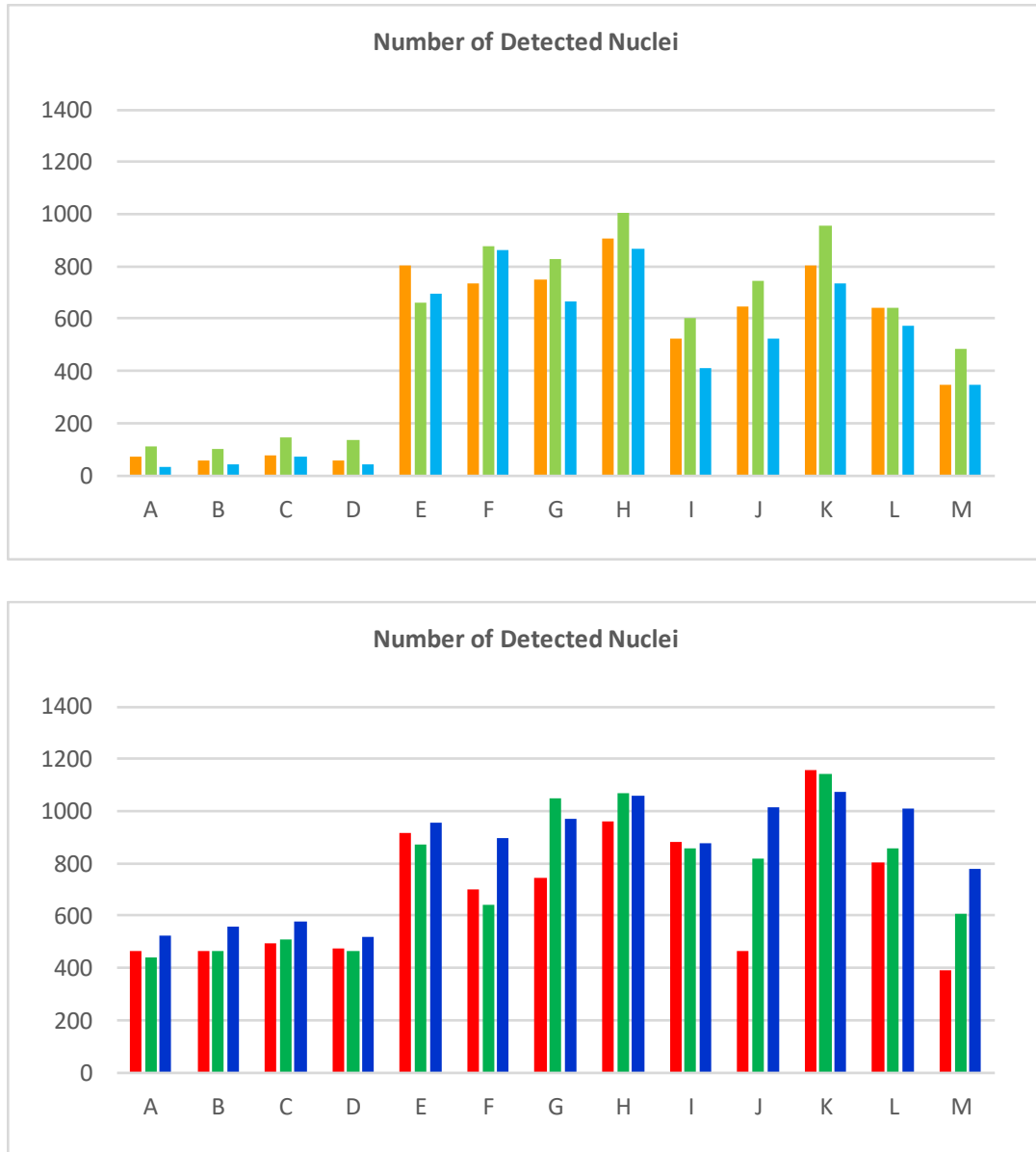


Figure 4.6 The number of nuclei detected on each ROI images, both on non-color corrected ROI images (above) and color corrected ROI images (below). In most cases, the number of detected nuclei improved by the application of stain-based color correction regardless of the WSS used to digitize the specimens [58].

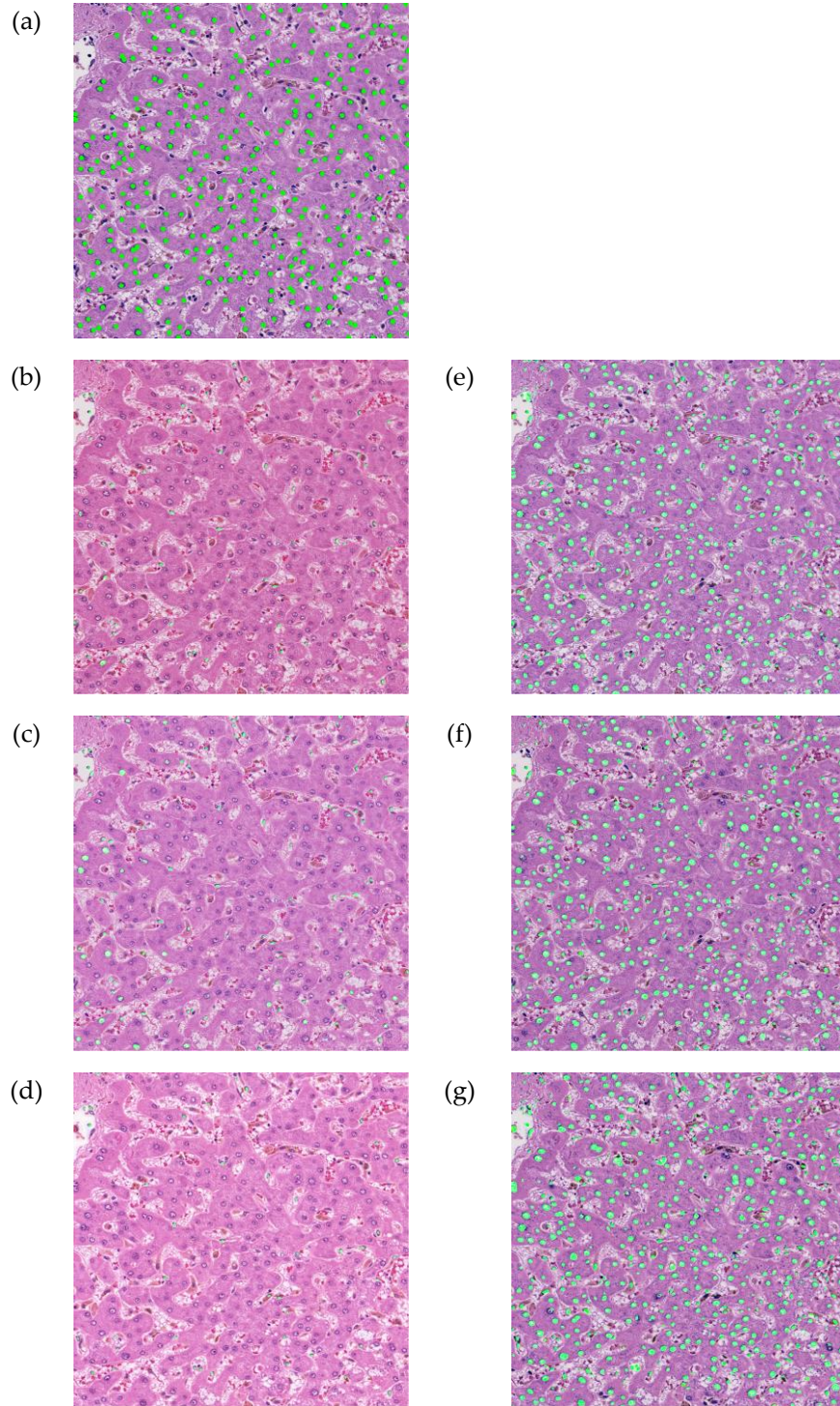


Figure 4.7 Comparison of nuclear detection results by the prototype system with manual detections (image B). (a) Manual detection. (b)-(d) Before color correction, (e)-(g) after color correction. (b) and (e) WSS-A, (c) and (f) WSS-B, and (d) and (g) WSS-C.

The detected nuclei are overlaid with green color [58],

Table 4.1 Comparison between the results of nuclei detection by the prototype system and manual detection [58].

Image B		
	Before color correction	After color correction
Manual	348	n/a
WSS-A	56	467
WSS-B	100	466
WSS-C	43	556
Image E		
	Before color correction	After color correction
Manual	839	n/a
WSS-A	804	916
WSS-B	659	871
WSS-C	695	958

4.2.4. Effect of stain-based color correction on quantified features

The effect of color correction on nuclei quantifications are presented in Table 4.2. Here, only the 50-percentile value on each nuclear feature are presented. The table shows the difference of quantified nuclear features between reference scanner and other, for the cases before and after color correction implementation. As mentioned in Subsection 4.2.1, quantified features from WSI-A are used as the reference. ΔAB means the quantification difference between images captured by WSS-B and WSS-A, while ΔAC means the quantification difference between images captured by WSS-C and WSS-A. The negative sign in ΔAB or ΔAC value means that the quantified feature on image captured by WSS-B or WSS-C is less than the value from the reference image. The value “n/a” means the feature cannot be calculated due to the number of detected nuclei is less than 100, as happened on the image A-D before color correction implementation.

Table 4.2 Value difference of 12 quantified nuclear features without and with color correction. Each feature is represented by its 50-percentile value [58].

Image	Area						Periphery					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	136.2	2.10	1.21	n/a	n/a	n/a	42.0	0.74	0.47
B	n/a	n/a	n/a	141.2	-1.49	-2.02	n/a	n/a	n/a	43.1	-0.08	0.25
C	n/a	n/a	n/a	123.0	4.24	5.92	n/a	n/a	n/a	40.4	1.10	0.82
D	n/a	n/a	n/a	134.1	0.61	-0.57	n/a	n/a	n/a	42.2	-0.05	0.49
E	79.5	5.51	-5.73	92.9	2.58	12.75	32.8	0.86	-1.33	35.2	0.42	2.45
F	86.0	2.29	-14.77	89.8	8.75	17.72	33.7	0.48	-2.74	34.6	1.31	2.71
G	126.3	0.71	-12.79	151.2	-5.04	10.89	41.9	-0.30	-2.44	45.3	-0.58	1.82
H	102.9	5.70	-1.44	118.7	3.54	20.64	37.5	0.66	-0.73	40.3	0.27	2.79
I	95.3	-5.04	-23.46	111.5	-6.49	3.02	35.8	-0.95	-4.17	38.4	-0.85	0.71
J	95.3	-1.33	-23.32	111.1	3.34	3.38	35.7	0.00	-3.81	39.1	0.18	0.07
K	150.0	3.44	-17.92	188.4	-12.18	-4.18	44.7	0.57	-2.77	50.2	-1.55	-0.49
L	90.1	4.58	-12.22	107.0	-2.47	2.06	34.2	1.00	-2.18	37.4	-0.64	0.27
M	81.7	5.44	-8.34	98.3	-0.19	1.83	33.0	1.12	-1.84	36.1	-0.08	0.21

Table 4.2. (continued).

Image	Circularity						Ratio of short & long axes lengths of elliptic fit					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	0.98	0.002	-0.004	n/a	n/a	n/a	0.80	0.002	-0.022
B	n/a	n/a	n/a	0.97	0.006	-0.002	n/a	n/a	n/a	0.77	0.025	-0.001
C	n/a	n/a	n/a	0.97	-0.001	-0.002	n/a	n/a	n/a	0.78	-0.004	-0.010
D	n/a	n/a	n/a	0.98	0.002	-0.008	n/a	n/a	n/a	0.79	0.012	-0.014
E	0.97	0.001	0.001	0.97	0.005	0.001	0.75	0.003	0.005	0.75	0.018	0.008
F	0.97	0.002	-0.012	0.97	0.006	0.003	0.77	0.010	-0.042	0.76	0.020	0.006
G	0.93	-0.006	0.001	0.93	-0.017	-0.011	0.67	-0.012	-0.003	0.68	-0.030	-0.020
H	0.95	0.002	0.000	0.95	0.003	0.002	0.71	0.007	0.001	0.70	0.003	0.008
I	0.97	-0.002	-0.019	0.97	0.000	0.001	0.76	0.003	-0.051	0.78	-0.005	0.008
J	0.96	0.005	-0.016	0.95	0.005	0.009	0.73	0.007	-0.044	0.71	0.019	0.032
K	0.95	-0.004	-0.002	0.92	0.009	0.013	0.71	-0.006	-0.001	0.68	-0.002	0.014
L	0.98	0.002	-0.008	0.98	0.004	0.001	0.79	0.011	-0.032	0.79	0.017	0.007
M	0.96	0.004	-0.002	0.96	0.006	0.007	0.74	0.016	-0.004	0.73	0.004	0.006

Table 4.2. (continued).

Image	Long axis lengths of elliptic fit						Short axis lengths of elliptic fit					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	7.499	0.078	0.100	n/a	n/a	n/a	5.813	-0.019	-0.182
B	n/a	n/a	n/a	7.678	0.035	0.123	n/a	n/a	n/a	5.716	0.050	0.063
C	n/a	n/a	n/a	7.328	0.104	0.194	n/a	n/a	n/a	5.402	-0.074	0.096
D	n/a	n/a	n/a	7.465	-0.036	0.166	n/a	n/a	n/a	5.588	0.103	-0.198
E	5.973	0.120	-0.235	6.327	0.045	0.437	4.319	0.114	-0.166	4.563	0.060	0.314
F	6.062	0.073	-0.447	6.251	0.204	0.361	4.531	0.098	-0.441	4.619	0.214	0.447
G	7.784	-0.142	-0.336	8.325	-0.004	0.363	4.964	-0.053	-0.247	5.507	-0.357	-0.073
H	6.939	0.065	-0.117	7.328	0.001	0.466	4.721	0.114	-0.055	4.987	0.161	0.456
I	6.509	-0.048	-0.635	6.957	-0.222	0.074	4.634	-0.142	-0.636	5.017	-0.096	0.051
J	6.554	-0.091	-0.646	7.097	-0.029	-0.011	4.609	-0.047	-0.643	4.897	0.044	0.139
K	8.198	-0.009	-0.603	9.114	-0.208	-0.134	5.722	-0.030	-0.364	6.310	-0.188	-0.074
L	6.063	0.125	-0.344	6.625	-0.195	0.006	4.724	0.146	-0.451	5.119	-0.021	0.043
M	6.016	0.160	-0.344	6.432	0.033	0.146	4.413	0.108	-0.287	4.764	0.002	0.074

Table 4.2. (continued).

Image	Cell contour complexity						Texture ASM of GLCM					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	51.95	8.79	-11.10	n/a	n/a	n/a	0.041	-0.002	-0.010
B	n/a	n/a	n/a	69.28	2.22	-24.60	n/a	n/a	n/a	0.039	-0.002	-0.008
C	n/a	n/a	n/a	55.70	6.98	-27.26	n/a	n/a	n/a	0.046	-0.005	-0.011
D	n/a	n/a	n/a	49.39	14.57	-11.84	n/a	n/a	n/a	0.036	-0.002	-0.006
E	3.38	2.89	-0.15	2.09	2.09	0.59	0.053	0.005	-0.008	0.068	-0.008	-0.005
F	3.25	0.07	-1.72	2.58	0.20	-1.08	0.047	0.009	-0.001	0.063	0.002	0.003
G	9.93	5.04	-1.16	9.14	2.69	0.14	0.043	-0.002	-0.008	0.052	0.001	0.004
H	4.97	-0.57	-0.49	3.34	0.30	0.86	0.051	0.024	-0.007	0.067	0.001	-0.002
I	11.85	7.44	-4.07	9.65	2.12	-1.48	0.040	0.001	0.001	0.044	0.006	0.000
J	6.30	-3.30	-3.50	10.23	-5.34	-8.11	0.039	0.012	0.006	0.043	0.011	0.007
K	14.19	8.14	1.68	20.51	-1.01	-10.80	0.039	0.002	-0.011	0.042	0.013	0.009
L	1.97	4.28	-0.19	3.06	1.06	-2.04	0.042	0.002	-0.001	0.052	0.006	-0.002
M	0.40	2.95	1.26	1.42	1.33	-1.12	0.048	0.001	-0.006	0.055	0.008	0.003

Table 4.2. (continued).

Image	Texture contrast of GLCM						Texture homogeneity of GLCM					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	333.9	8.83	39.47	n/a	n/a	n/a	0.330	-0.013	0.011
B	n/a	n/a	n/a	357.7	3.35	3.22	n/a	n/a	n/a	0.325	-0.007	0.022
C	n/a	n/a	n/a	377.0	-1.03	1.52	n/a	n/a	n/a	0.335	-0.008	0.019
D	n/a	n/a	n/a	338.1	4.57	17.82	n/a	n/a	n/a	0.320	-0.007	0.026
E	179.6	44.27	37.30	357.2	-12.11	-33.16	0.464	-0.008	-0.030	0.467	-0.010	0.004
F	321.7	18.73	18.64	361.4	9.21	3.67	0.433	-0.001	-0.017	0.467	-0.013	0.017
G	187.9	34.10	22.66	253.2	10.49	-1.23	0.430	-0.033	-0.038	0.443	-0.023	0.009
H	200.0	22.07	22.66	286.6	-6.26	-7.83	0.467	0.017	-0.044	0.489	-0.009	-0.001
I	308.3	-1.82	-4.99	336.1	15.74	-47.33	0.367	-0.018	-0.012	0.399	-0.025	-0.038
J	311.9	-3.64	-3.42	335.4	3.74	-14.34	0.396	0.019	0.009	0.410	0.024	0.018
K	179.5	22.14	-2.97	193.9	42.59	10.57	0.434	-0.026	-0.039	0.443	0.005	0.007
L	325.3	8.64	-4.57	344.6	28.44	-9.62	0.422	-0.038	-0.030	0.458	-0.005	-0.035
M	360.1	0.98	2.92	388.5	1.72	-21.55	0.424	-0.035	-0.035	0.455	-0.005	-0.023

Table 4.2. (continued).

Image	Texture entropy of GLCM						Nuclei Density					
	w/o color correction			with color correction			w/o color correction			with color correction		
	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC	WSI-A	ΔAB	ΔAC
A	n/a	n/a	n/a	4.55	-0.005	0.024	n/a	n/a	n/a	5	0	1
B	n/a	n/a	n/a	4.54	0.056	0.097	n/a	n/a	n/a	5	0	1
C	n/a	n/a	n/a	4.45	0.018	0.080	n/a	n/a	n/a	5	0	1
D	n/a	n/a	n/a	4.60	0.004	-0.065	n/a	n/a	n/a	5	0	1
E	3.88	-0.015	0.148	3.89	0.047	0.082	8	0	-1	10	0	0
F	4.03	0.004	0.106	3.88	0.036	-0.004	9	1	0	9	1	2
G	4.15	0.175	0.197	4.13	0.103	0.005	8	1	-1	8	3	3
H	3.93	-0.137	0.227	3.84	0.032	0.083	10	1	0	11	1	1
I	4.36	0.024	-0.024	4.22	0.048	0.180	5	1	-1	9	0	0
J	4.22	-0.011	-0.087	4.18	-0.096	0.026	8	1	-1	8	2	3
K	4.22	0.093	0.245	4.19	-0.100	0.082	10	1	-1	13	0	-1
L	4.13	0.141	0.106	4.01	-0.067	0.190	8	0	-2	10	0	1
M	4.16	0.085	0.107	4.08	-0.086	0.060	6	1	-1	8	0	2

Figure 4.8 presents the bar charts of three nuclear features, represented by the 50-percentile values, where two features are related to nuclei's morphology (area and peripheral), and one feature is related to nuclei's texture (contrast of gray-level co-occurrence matrix (GLCM)). The WSI-A1, WSI-B1, and WSI-C1 labels refer to non-color corrected ROI images captured by WSS-A, WSS-B, and WSS-C respectively (Figures 4.8 (a), (c), (e), (g), (i), (k), (m), (o), (q), (s), (u), and (w)). On the other hand, WSI-A2, WSI-B2, and WSI-C2 labels are for the color-corrected version of the ROI images. The graph shows that in considerable cases, the difference between the features calculated from different scanners are reduced after color correcting the ROI images, such as in the case of image J, K, and M on nuclei's area and peripheral features, thus improving the quantified nuclear features. On the other hand, the quantified features become worse as shown in the case of image I, K, L, and M in the texture contrast of GLCM feature. The quantified features are neither improved nor getting worse with the color correction some other cases. Analyzing further on the ROI images shows that even though the color differences were reduced, some variations can still be observed on the nuclei. These results show that the result of stain-based color correction is more effective on certain features, such as features related to nuclei's morphology, and can be less effective on texture-related features.

On the other hand, variations on image sharpness may also cause the variations on the nuclei quantification, especially on the texture related features. These phenomena can be observed on images shown in Figure 4.9. Images on the first line show specific regions on image I after color correction, while images on the second line show the luminance component of the images in the first line. The two graphs display the luminance profile of labeled nuclei along the green line from images captured by WSS-A, WSS-B, and WSS-C. On the luminance profile graph, about 10 to 11 pixels on the left and right side of the nuclei are included to check the sharpness of the images. The slopes on the blue graphs on the nuclei's border are smaller compared to the red and green graphs, confirming that the sharpness is different on the images. Sharpness difference can be caused by the imaging system in the scanning device, and it can lead to different quantification results on the nuclei's features related to texture.

- WSI-A1 : Non-color corrected ROI images scanned by WSS-A
- WSI-B1 : Non-color corrected ROI images scanned by WSS-B
- WSI-C1 : Non-color corrected ROI images scanned by WSS-C
- WSI-A2 : Color corrected ROI images scanned by WSS-A
- WSI-B2 : Color corrected ROI images scanned by WSS-B
- WSI-C2 : Color corrected ROI images scanned by WSS-C



Figure 4.8 The 50-percentile values of 12 nuclear features on each case [58].

- WSI-A1 : Non-color corrected ROI images scanned by WSS-A
- WSI-B1 : Non-color corrected ROI images scanned by WSS-B
- WSI-C1 : Non-color corrected ROI images scanned by WSS-C
- WSI-A2 : Color corrected ROI images scanned by WSS-A
- WSI-B2 : Color corrected ROI images scanned by WSS-B
- WSI-C2 : Color corrected ROI images scanned by WSS-C



Figure 4.8. (continued).

- WSI-A1 : Non-color corrected ROI images scanned by WSS-A
- WSI-B1 : Non-color corrected ROI images scanned by WSS-B
- WSI-C1 : Non-color corrected ROI images scanned by WSS-C
- WSI-A2 : Color corrected ROI images scanned by WSS-A
- WSI-B2 : Color corrected ROI images scanned by WSS-B
- WSI-C2 : Color corrected ROI images scanned by WSS-C



Figure 4.8. (continued).

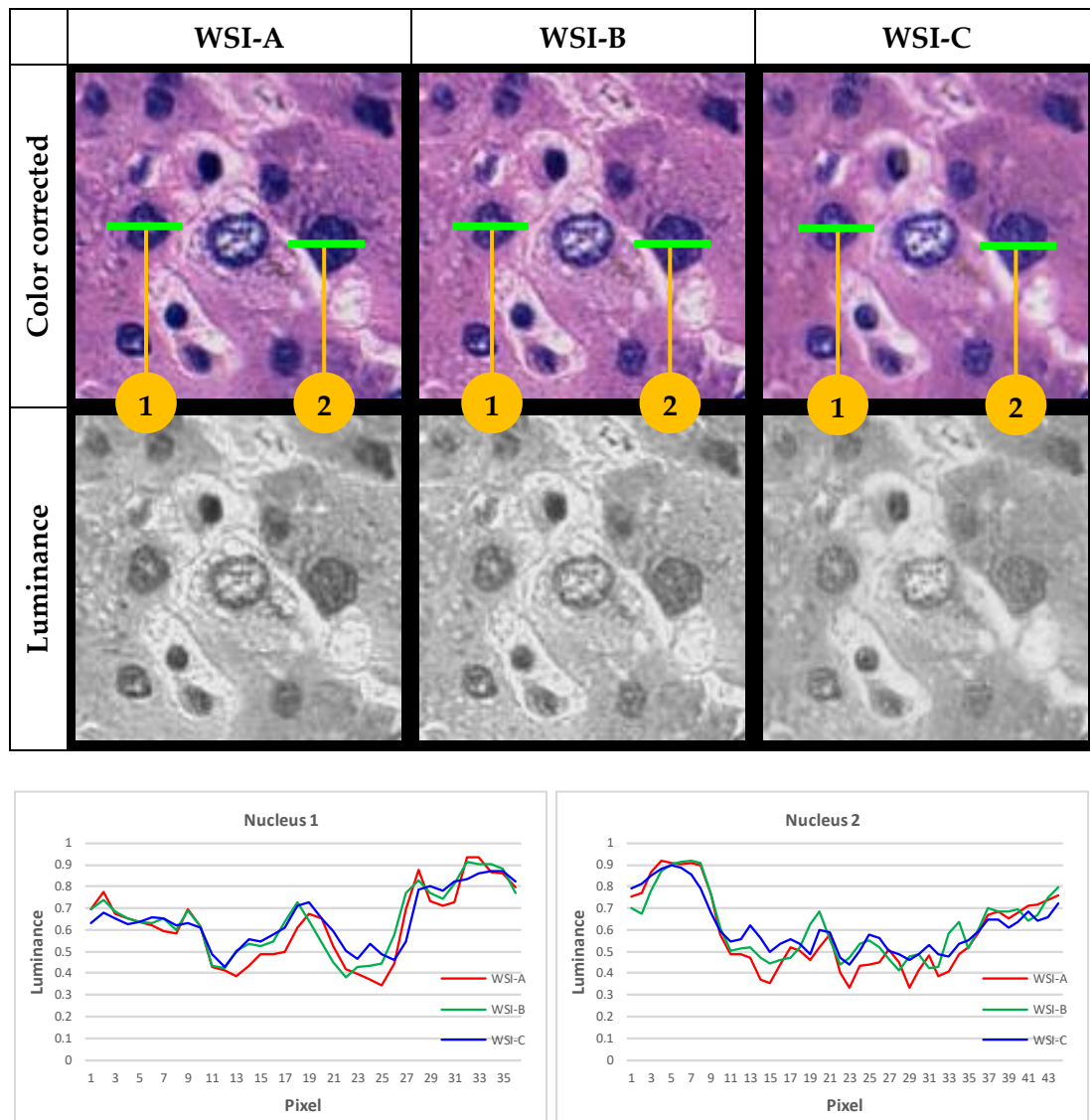


Figure 4.9 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs [58].

4.3 Effect of device-based color correction on histopathological images acquired by different WSSs

This section focuses on the investigation of the effect of device-based color correction on WSIs captured by different devices. The evaluation was performed by comparing the color distribution of the tissue components before and after color correction. The feature quantification was not implemented on device-based color-corrected images since the comparison on quantified features was not performed since the device-based color correction will not be able to compensate the color difference due to staining process. Meanwhile, the feature quantification on the prototype system will only be effective if the stain-based color difference is corrected.

4.3.1. Experiment workflow

This experiment utilized the same three liver specimens and three scanner devices as the experiments in Section 4.2. Similar processes were performed on the WSIs. Only three ROI images were used in this experiment, which were images A, E and I (one ROI image from each sample).

Two color correction matrices were derived to characterize WSS-B and WSS-C. The color correction matrices were derived from a color chart slide which contain 9 color patches. The color chart slide has the same specification with the one introduced by Bautista et al. [55]. The slide was digitized using the three WSSs and the digitized slide were then converted to TIF files as shown in Figure 4.10 below.

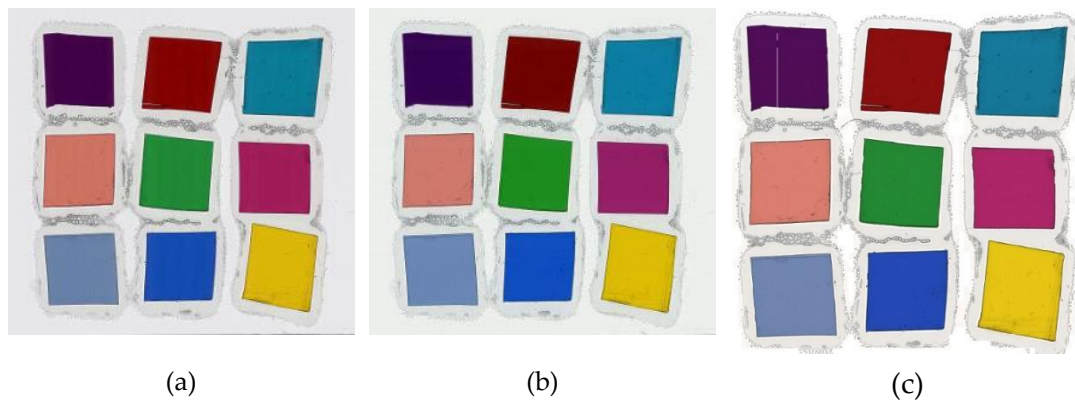


Figure 4.10 Digitized color chart slide by (a) WSS-A, (b) WSS-B, and (c) WSS-C.

A region of 100×100 pixels was selected for each patch, and a scanned white region in between the patches to represent white color as shown in Figure 4.11. Thus, 10 colors were selected, labeled as blue, cyan, green, magenta, pink, purple, red, sky, white, and yellow. The average value on each R, G, and B channels were then calculated for each color. The color correction matrices for image captured by WSS-B and WSS-C were then derived by using pure quadratic model. M_B is the color correction matrix for image captured by WSS-B while M_C is the color correction matrix for image captured by WSS-C. The color correction matrices are shown below.

$$M_B = \begin{bmatrix} 1.2151 & -0.0877 & 0.0518 & -0.0009 & 0.0003 & -0.0003 \\ -0.0008 & 0.9871 & -0.0218 & -0.0000 & 0.0000 & 0.0001 \\ 0.0600 & -0.0154 & 0.9176 & -0.0006 & 0.0009 & -0.0002 \end{bmatrix}$$

$$M_C = \begin{bmatrix} 0.9018 & 0.0543 & -0.0965 & 0.0003 & -0.0003 & 0.0003 \\ -0.1518 & 1.1559 & -0.0498 & 0.0004 & -0.0007 & 0.0003 \\ -0.1675 & 0.1157 & 0.9075 & 0.0005 & -0.0004 & 0.0004 \end{bmatrix}$$

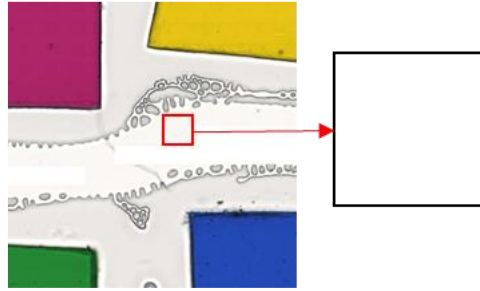


Figure 4.11 Obtaining white color on from the digitized color chart slide. A region of 100×100 pixels on scanned white area between the patches was selected to represent the white color for the color correction matrix.

The uncorrected color chart slide and histopathological images were compared with the color corrected version to confirm the effect of device-based color correction on WSIs captured by different devices. In analyzing the histopathological images, the six tissue components were compared by using the same procedure with the experiment in Section 4.2. Figure 4.12 describes the overall process of this experiment

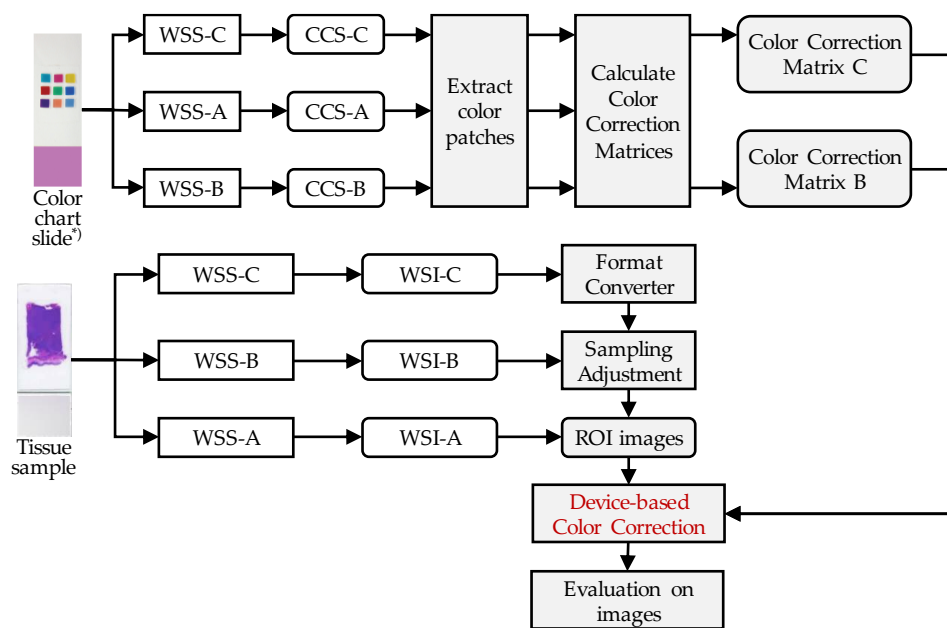


Figure 4.12 Workflow of experiment on device-based color correction. A color chart slide and three WSSs were used to derive color correction matrices B and C. The color correction matrices were then used to correct the color on images captured by WSS-B and WSS-C. Three ROI images of 3 liver specimens were selected for this experiment.

Image of the color chart slide is taken from [55].

4.3.2. Effect of device-based color correction on digitized color chart slide

Figure 4.13 shows the effect of device-based color correction on each color patch in the digitized color chart slide. The mean of RGB values for each color are shown below the image. The color on images captured by WSS-B and WSS-C after color correction becomes similar to the color on images captured by WSS-A. Figure 4.14–Figure 4.16 show the projection of cluster plots of the mean RGB values for each color on the L-a, L-b, and a-b chromaticity axis of CIELAB color space, respectively. The number 1 and 2 after the WSS-B and WSS-C refer to the uncorrected and color corrected version of the image. All graphs show that the implementation of device-based color correction makes the color distribution on the color chart slide images obtained by different WSSs becomes more similar. The WSI-A1, WSI-B1, and WSI-C1 labels refer to the original version of ROI images captured by WSS-A, WSS-B, and WSS-C, respectively. On the other hand, WSI-A2, WSI-B2, and WSI-C2 labels are for the device-based color corrected version.

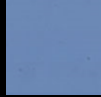
Color	WSS-A (Ref)	WSS-B		WSS-C	
		w/o correction	with correction	w/o correction	with correction
Blue	 (19.3, 74.8, 175.3)	 (15.5, 76.7, 192.4)	 (12.5, 75.1, 174.0)	 (36.4, 73.3, 180.9)	 (28.3, 76.8, 177.9)
Cyan	 (16.7, 124.5, 150.6)	 (13.3, 125.7, 162.4)	 (10.3, 122.7, 156.5)	 (22.9, 119.2, 152.6)	 (14.9, 124.2, 152.6)
Green	 (30.1, 138.0, 53.8)	 (27.1, 139.9, 30.5)	 (27.1, 137.9, 44.5)	 (36.3, 136.5, 51.9)	 (30.6, 137.9, 51.0)
Magenta	 (160.7, 29.8, 90.1)	 (147.0, 29.6, 100.1)	 (159.0, 27.6, 86.0)	 (172.4, 38.3, 102.9)	 (159.4, 27.3, 87.9)
Pink	 (206.3, 114.9, 99.2)	 (206.3, 121.9, 110.8)	 (208.3, 118.9, 97.6)	 (219.6, 126.3, 112.5)	 (207.6, 119.3, 102.5)
Purple	 (66.3, 1.2, 70.5)	 (54.9, 4.2, 81.1)	 (65.9, 3.2, 74.2)	 (80.6, 13.7, 87.5)	 (69.5, 3.7, 73.5)
Red	 (129.6, 1.8, 0.5)	 (110.3, 1.9, 0.0)	 (123.3, 1.9, 0.0)	 (141.2, 14.9, 15.1)	 (132.5, 2.9, 2.1)
Sky	 (103.7, 129.7, 166.9)	 (107.2, 135.7, 180.7)	 (113.2, 133.0, 173.3)	 (112.1, 134.2, 177.1)	 (99.1, 131.2, 169.1)
White	 (232.0, 231.8, 233.7)	 (235.4, 237.4, 235.6)	 (227.6, 234.4, 233.0)	 (250.7, 246.9, 243.6)	 (234.7, 235.5, 238.6)
Yellow	 (219.4, 184.7, 8.7)	 (219.5, 187.4, 5.2)	 (217.6, 184.4, 17.7)	 (231.0, 197.1, 20.7)	 (221.6, 186.1, 14.3)

Figure 4.13 The effect of device-based color correction on the ten color patches from the digitized color chart slide. The color difference on the ROI images were reduced by color correction.

- ★ WSI-A1 : ROI image captured by WSS-A
- WSI-B1 : ROI image captured by WSS-B ● WSI-B2 : Color-corrected ROI image acquired by WSS-B
- △ WSI-C1 : ROI image captured by WSS-C ▲ WSI-C2 : Color-corrected ROI image acquired by WSS-C

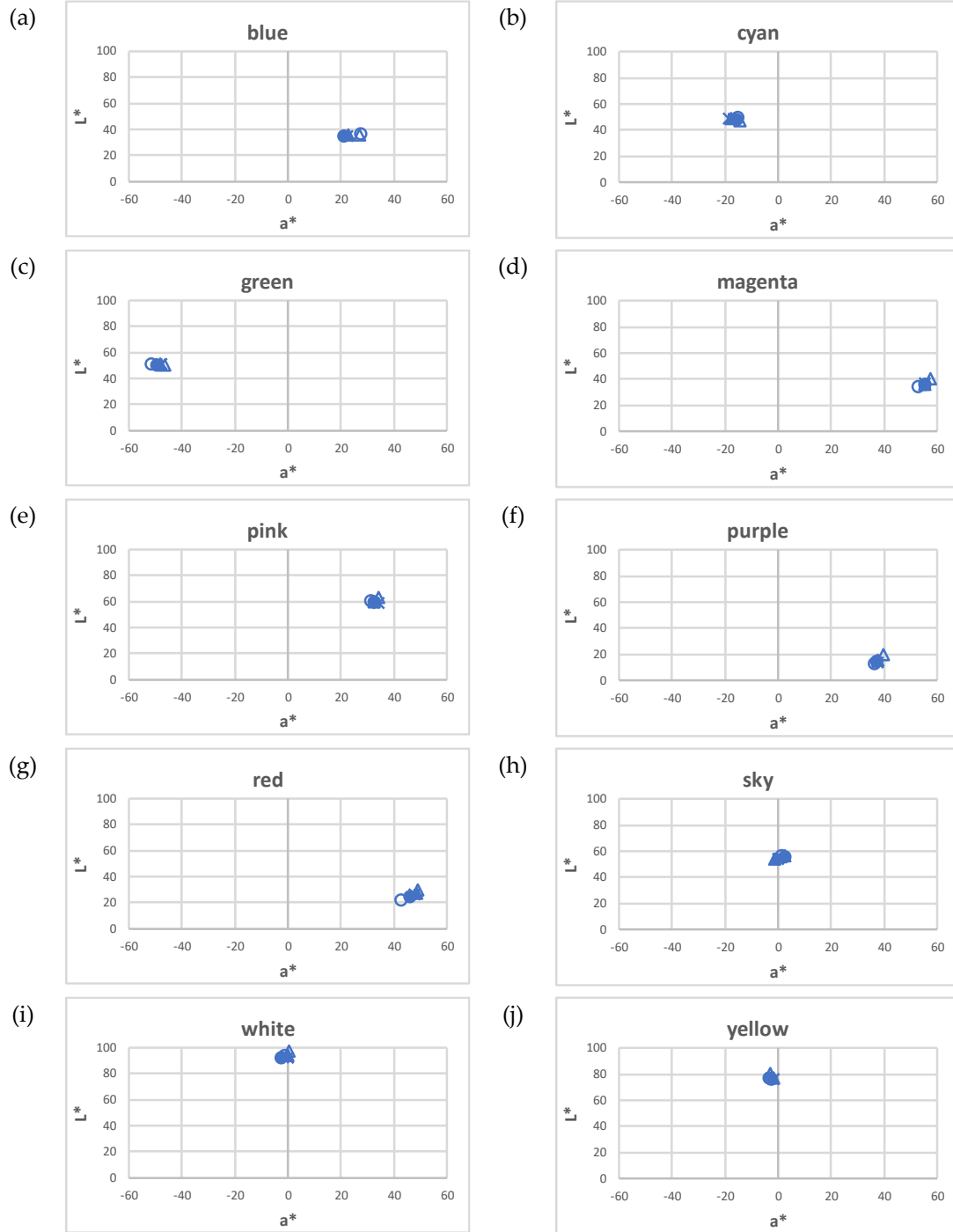


Figure 4.14 The projection of cluster plots of mean RGB values on each color patch on the L-a chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.

- ★ WSI-A1 : ROI image captured by WSS-A
 ○ WSI-B1 : ROI image captured by WSS-B ● WSI-B2 : Color-corrected ROI image acquired by WSS-B
 △ WSI-C1 : ROI image captured by WSS-C ▲ WSI-C2 : Color-corrected ROI image acquired by WSS-C

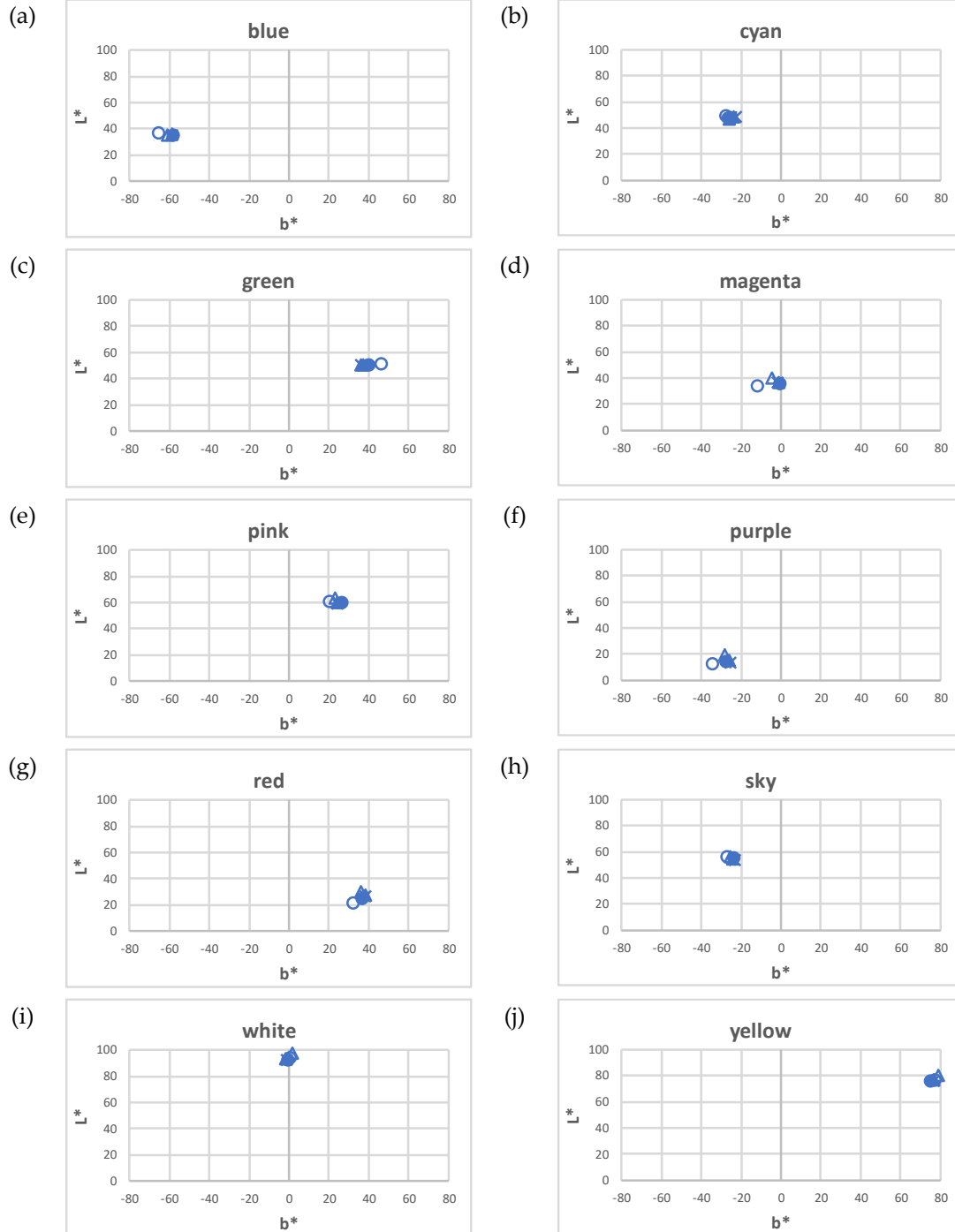


Figure 4.15 The projection of cluster plots of mean RGB values on each color patch on the L-b chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.

- ★ WSI-A1 : ROI image captured by WSS-A
- WSI-B1 : ROI image captured by WSS-B ● WSI-B2 : Color-corrected ROI image acquired by WSS-B
- △ WSI-C1 : ROI image captured by WSS-C ▲ WSI-C2 : Color-corrected ROI image acquired by WSS-C

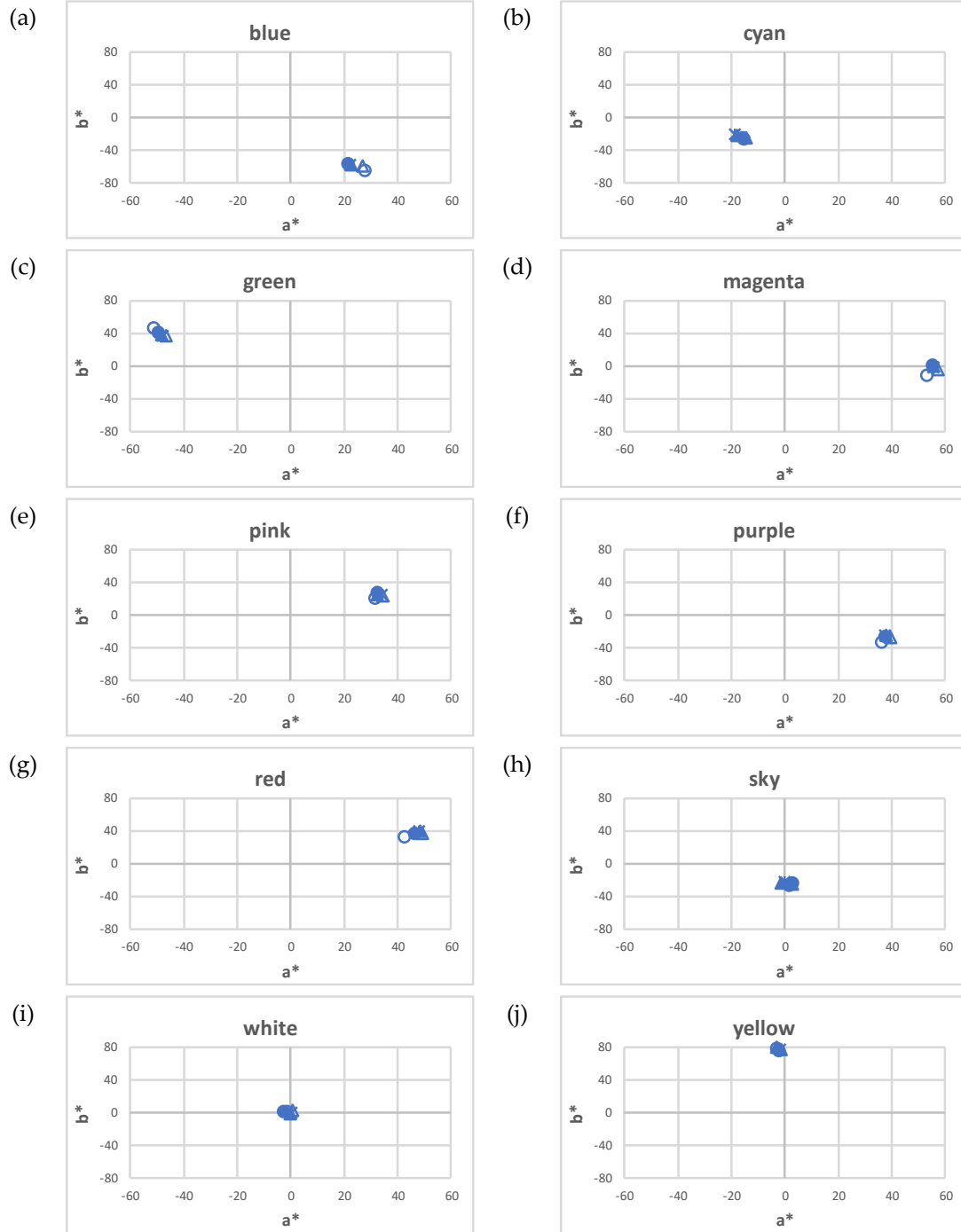


Figure 4.16 The projection of cluster plots of mean RGB values on each color patch on the a - b chromaticity axes of CIELAB color space. The number 1 and 2 after WSS-B and WSS-C refer to uncorrected and color corrected version of the WSI.

4.3.3. Effect of stain-based color correction on tissue components

Figure 4.17 shows the effect of device-based color correction on each ROI image. There are changes on the color after color correction but the colors have not been similar to the color in the reference images, unlike the effect on the color chart slide image. Nevertheless, after implementing the color correction, the blue color on ROI images extracted from WSI-B is less strong, while the color on ROI images extracted from WSI-C become less yellowish. Figure 4.18 shows the projection of cluster plots of six tissue components on the a-b chromaticity axes of CIELAB color space, where each tissue components were represented with 30 data samples, similar to the experiment in stain-based color correction. The result shows that there are different effects when the color correction technique is implemented on digitized color chart slide and the histopathological images. The graphs also show that the circles are not as compact as when stain-based color correction was implemented (Figure 4.5). This result indicates that the device-based color correction is less effective when applied on the histopathological images compared to the stain-based color correction.

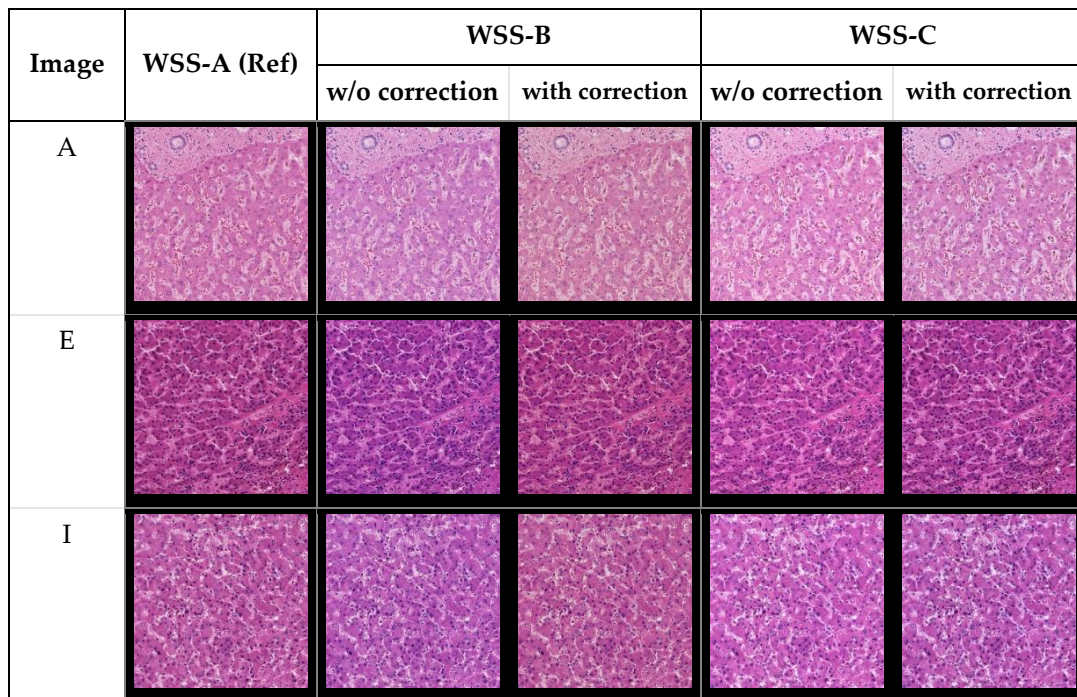


Figure 4.17 The effect of device-based color correction on the three sets of three ROI images. The colors on the color-corrected images changes although the colors have yet to become similar to the reference images.

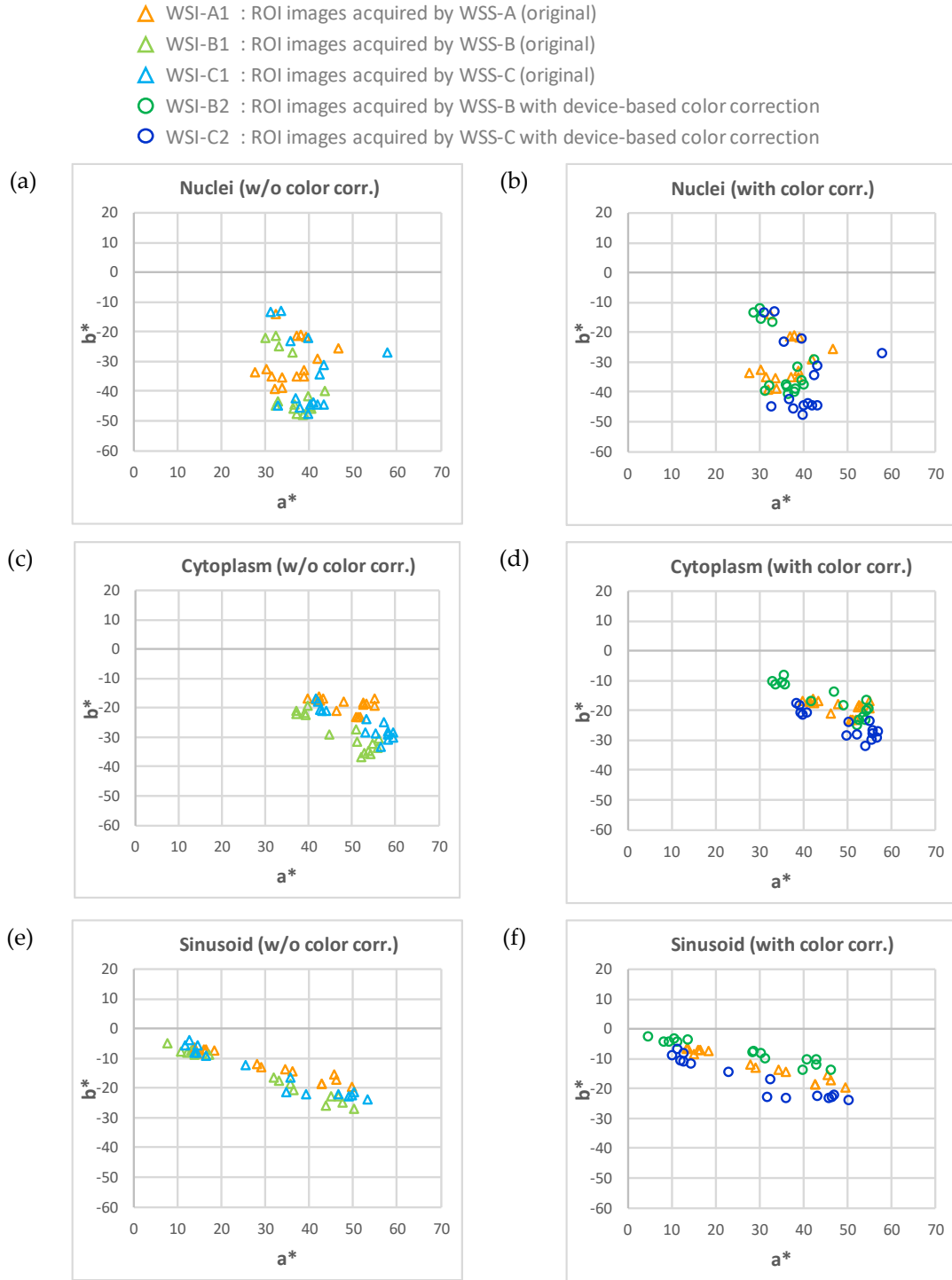


Figure 4.18 The projection of cluster plots of six tissue components (nuclei, cytoplasm, sinusoid, lymphocyte, fiber, red blood cell) on the a - b chromaticity axes of CIELAB color space. The number 1 and 2 after the WSS type refer to uncorrected and color corrected version of the WSI.

- △ WSI-A1 : ROI images acquired by WSS-A (original)
- △ WSI-B1 : ROI images acquired by WSS-B (original)
- △ WSI-C1 : ROI images acquired by WSS-C (original)
- WSI-B2 : ROI images acquired by WSS-B with device-based color correction
- WSI-C2 : ROI images acquired by WSS-C with device-based color correction

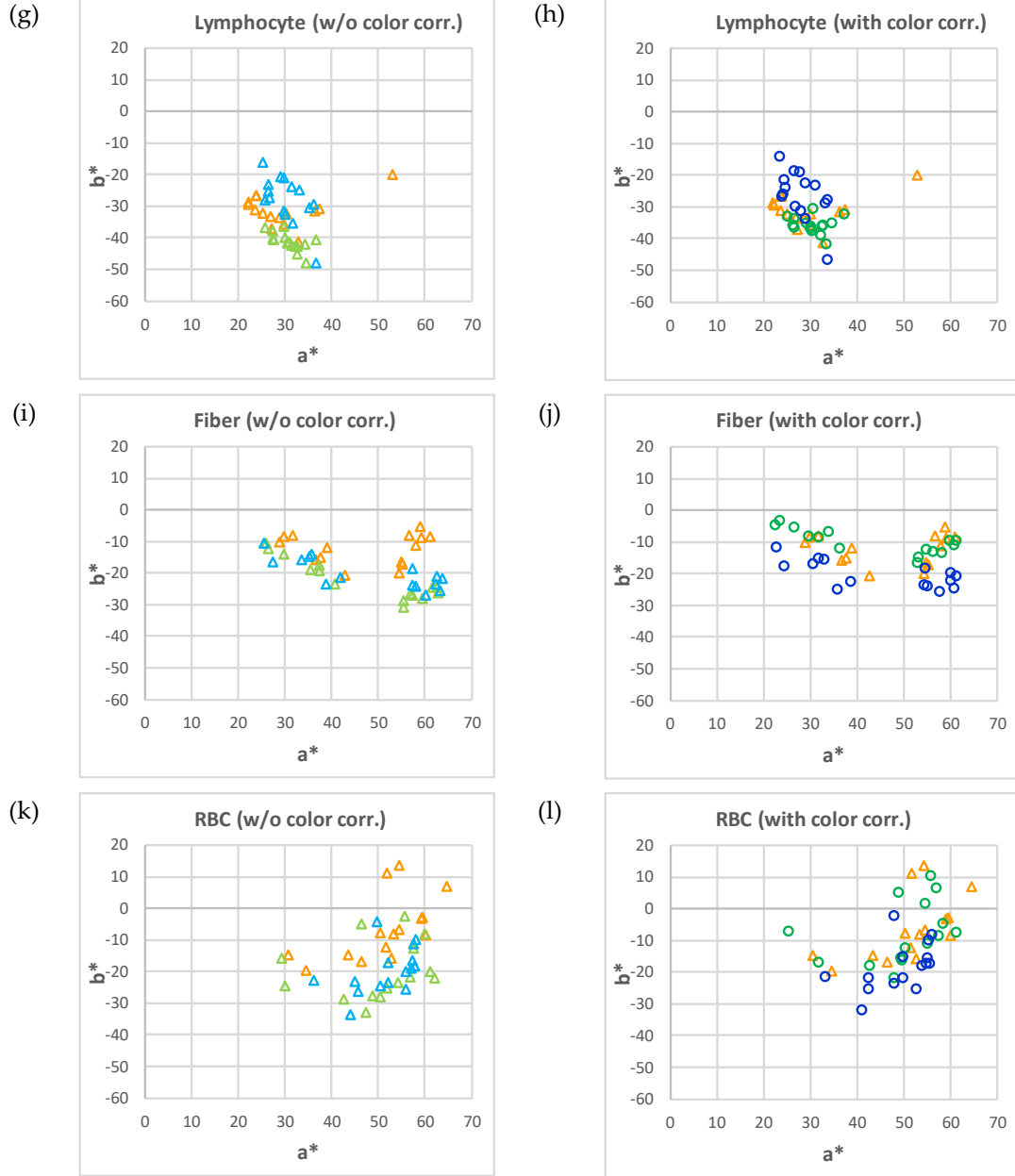


Figure 4.18. (continued).

Table 4.3 shows the average of color difference (ΔE^*) between image captured by WSS-B and WSS-C with the reference image, before and after color correction. The color difference ΔE^* is calculated by using Equation (4.1), where the (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) are the color components in the CIELAB for each image captured by the first and second device.

$$\Delta E^* = \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2} \quad (4.1)$$

After device-based color correction, the color differences on the tissue components were reduced except on sinusoid of images obtained by WSS-B.

Table 4.3 Effect of device-based color correction on the tissue components.

Tissue Component	w/o color correction		with color correction	
	$\Delta E_{WSSA-WSSB}$	$\Delta E_{WSSA-WSSC}$	$\Delta E_{WSSA-WSSB}$	$\Delta E_{WSSA-WSSC}$
Nuclei	11.12	10.29	7.10	8.13
Cytoplasm	10.85	9.85	5.92	7.43
Sinusoid	5.84	8.93	7.00	7.38
Lymphocytes	12.93	11.55	8.89	14.00
Fiber	10.62	15.25	6.47	10.46
Red Blood Cell	17.09	15.18	8.25	13.43

4.4. Effect of combination of device- and stain-based color correction on histopathological images acquired by different WSSs

This section focuses on the investigation of the combination of device- and stain-based color correction on WSIs captured by different devices and its effect on the histopathological images and quantified features.

4.4.1. Experiment workflow

The experiment workflow was basically a combination of experiment in Subsections 4.2.1 and 4.3.1. Two ROI images captured by WSS-A, WSS-B, and WSS-C from sample 1 and 2 were used in this experiment, which were images A and E. The ROI images obtained by the WSS-B and WSS-C were corrected by using the color correction matrices introduced in Subsection 4.3.1. The device-based color-corrected images were then corrected by using the stain-based color correction method in Section 4.2. Note that the parameters for the stain-based color correction was adjusted since the color had been transformed by the device-based color correction. Figure 4.19 illustrated the workflow of the experiment.

The evaluation on the method was performed by comparing the images captured by WSS-B and WSS-C with the reference image. In this case, stained-base color-corrected ROI images acquired by WSS-A were used as the reference images. Thus, in each case, a set of five images were prepared as shown on Table 4.4.

Table 4.4 A set of five images for each case in the experiment to evaluate the combination of device- and stain-based color correction.

Image ID	WSS	Color correction	
		Stain-based	Device-based
WSI-A1	A (ref.)	Yes	No
WSI-B1	B	Yes	No
WSI-C1	C	Yes	Yes
WSI-B2	A	Yes	Yes
WSI-C2	B	Yes	Yes

For each case, three evaluations were performed; evaluation on the number of nuclei detected, quantified features, and the intensity profiles of several objects (nuclei and lymphocytes) on the images.

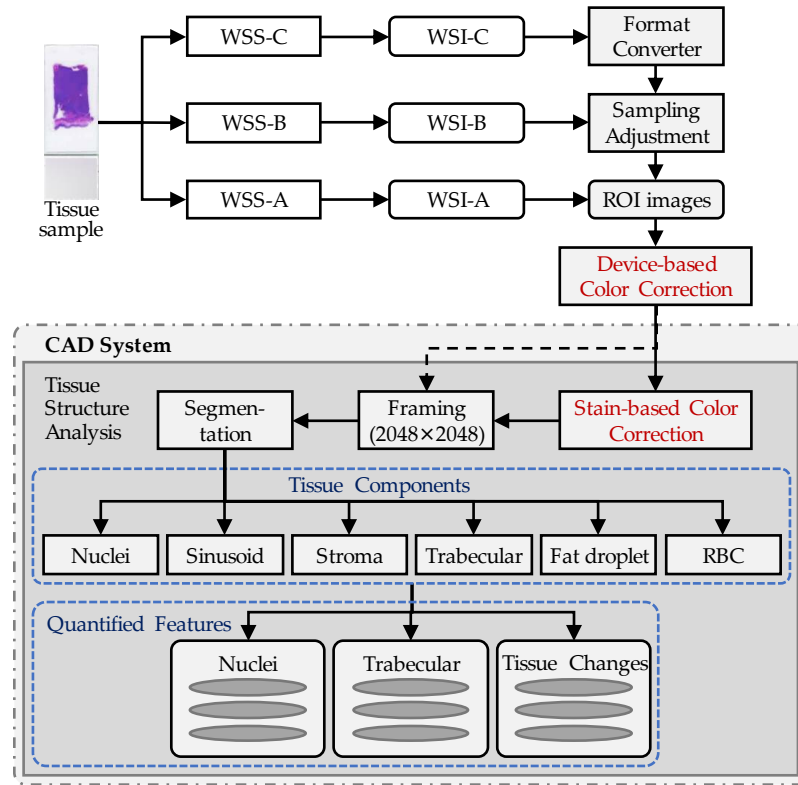


Figure 4.19 Workflow of experiment on combination of device- and stain-based color correction. Two ROI images from two liver specimens were used, where each specimen was scanned by WSS-A, WSS-B, and WSS-C. Device- and stain-based color corrections were performed sequentially on the ROI images.

4.4.2. Effect of combination of device- and stain-based color correction on nuclei detection

The effect of device- and stain-based color correction on nuclei detection is shown in Figure 4.20, where the effect is different based on the cases. On images captured by WSS-B, different effect can be observed where the number of nuclei detected was increased in Image A but decreased in Image E. Meanwhile, there are less changes on the number of nuclei detected on images captured by WSS-C. Yet, the effect is opposing the image

captured by WSS-B. Table 4.5 shows the differences of the number of nuclei detected on each image case. $\Delta B1A1$ and $\Delta C1A1$ means the result difference between stain-based color corrected images captured by WSS-B and WSS-C, respectively, with the stain-based color-corrected images captured by WSS-A. Meanwhile, $\Delta B2A1$ and $\Delta C2A1$ means the result difference between combination of device- and stain-based color corrected images captured by WSS-B and WSS-C, respectively, with the stain-based color-corrected images captured by WSS-A. The symbol “ $\downarrow$ ” means the difference between combined color-corrected image and the stain-based color-corrected image is reduced, the symbol “ $\uparrow$ ” means the difference between combined color-corrected image and the stain-based color-corrected image is increased, while “n/a” means there is no changes on the result. The results show that the effect of combination of device- and stain-based color correction on nuclei detection are vary based on the cases.

- WSI-A1 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B1 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C1 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-C

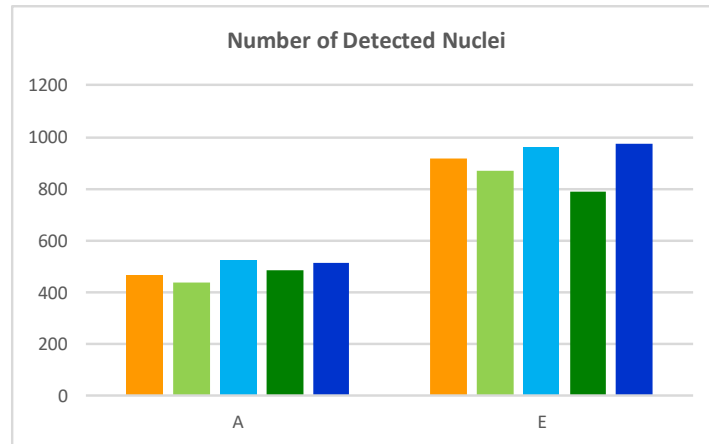


Figure 4.20 Comparison on the number of nuclei detected on image before and after sharpness correction.

Table 4.5 Differences on the number of nuclei detected between images before and after sharpness correction.

Image	WSI-B			WSI-C		
	$\Delta B1A1$	$\Delta B2A1$	Diff.	$\Delta C1A1$	$\Delta C2A1$	Diff.
A	26	19	↑	56	50	↓
E	45	129	↓	42	57	↑

4.4.3. Effect of combination of device- and stain-based color correction on quantified features

The effect of combination of color corrections on nuclei quantifications are presented in Figure 4.21. The 50-percentile values on each nuclear feature extracted from stain-based color-corrected images are compared with images corrected with a combination of device- and stain-based color correction, where stain-based color corrected images obtained by WSS-A is used as the reference images. The WSI-A1, WSI-B1, and WSI-C1 labels refer to stain-based color-corrected ROI images captured by WSS-A, WSS-B, and WSS-C, respectively. On the other hand, WSI-A2, WSI-B2, and WSI-C2 labels are for images with a combination of device- and stain-based color-correction, captured by WSS-A, WSS-B, and WSS-C, respectively. The graphs show that the effect on the combination of color corrections is vary and depends on the cases. However, less changes can be found on morphological-related nuclear features such as area and periphery. Table 4.6 and Table 4.7 show the summaries of differences on the quantified nuclear features on image A and E, respectively. The combination of device- and stain-based color correction gives positive result in most nuclear features extracted from image A that was scanned by WSS-B.

- WSI-A1 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B1 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C1 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-C

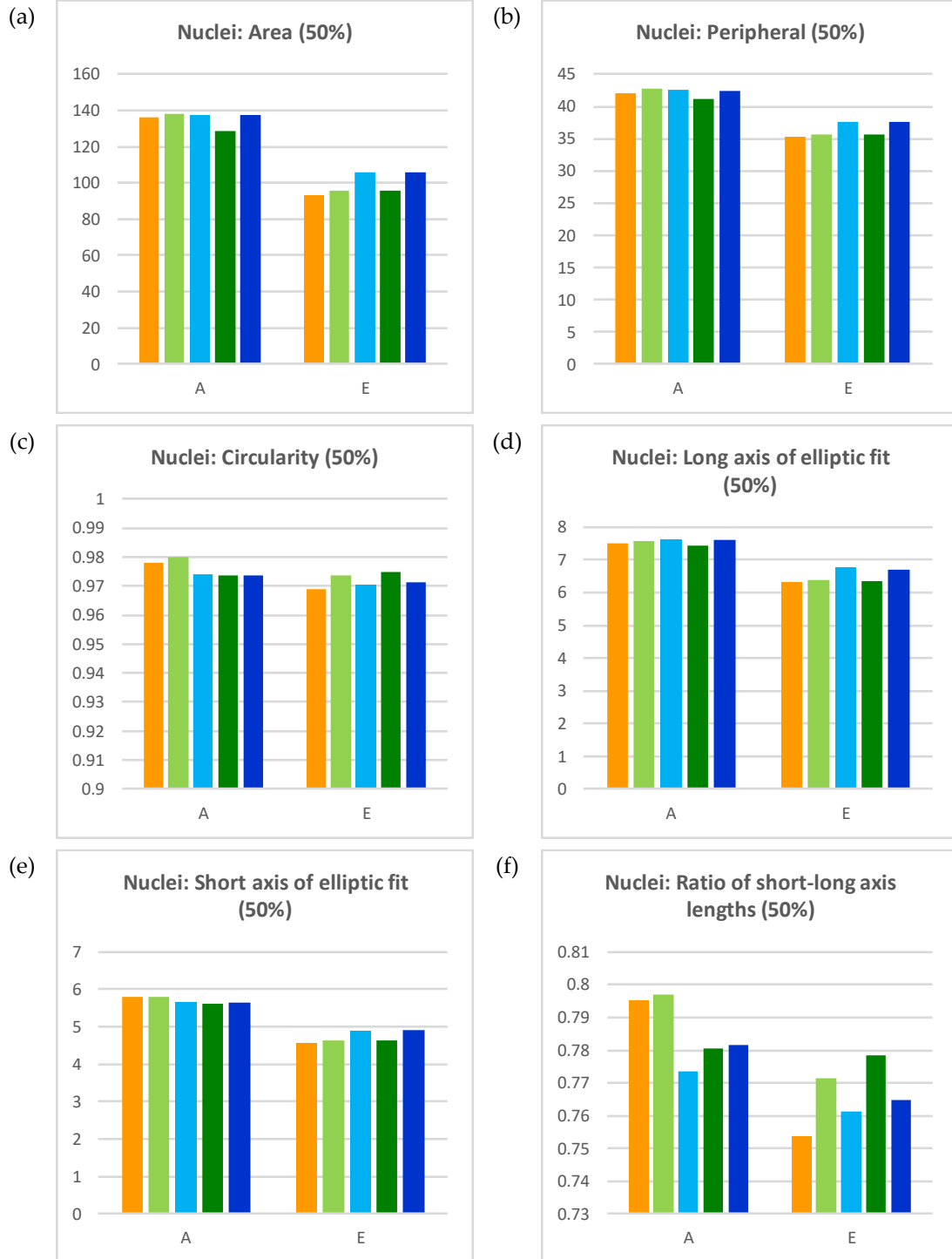


Figure 4.21 The 50-percentile values of 12 nuclear features on each case.

- WSI-A1 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B1 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C1 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Combination of device- and stain-based color corrected ROI images acquired by WSS-C

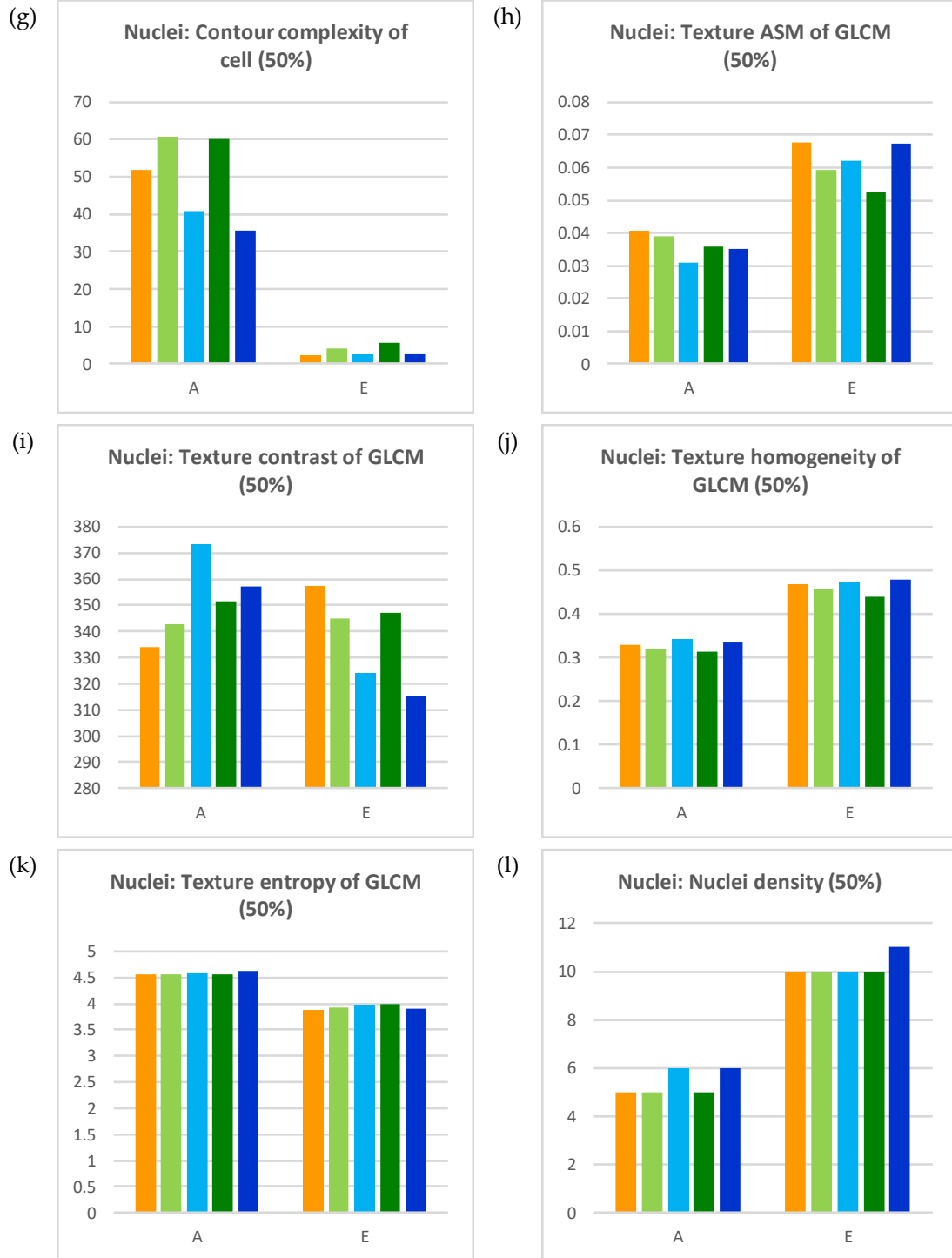


Figure 4.21. (continued).

Table 4.6 Differences on the nuclear feature extracted from stained-based color-corrected image A without and with device-based color correction where the nuclear feature on reference image (WSI-A).

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B1A1$	$\Delta B2A1$	Diff.	$\Delta C1A1$	$\Delta C2A1$	Diff.
1	Area	2.098	-7.888	↓	1.211	1.300	↑
2	Peripheral	0.741	-0.820	↓	0.466	0.418	↓
3	Circularity	0.002	-0.004	↓	-0.004	-0.004	↓
4	Long axis of elliptic fit	0.078	-0.065	↓	0.100	0.117	↑
5	Short axis of elliptic fit	-0.019	-0.212	↓	-0.182	-0.174	↑
6	Ratio of short-long axis lengths	0.002	-0.015	↓	-0.022	-0.014	↑
7	Contour complexity of cell	8.790	8.227	↓	-11.100	-16.432	↓
8	Texture ASM of GLCM	-0.002	-0.005	↓	-0.010	-0.006	↑
9	Texture contrast of GLCM	8.826	17.468	↑	39.468	23.341	↓
10	Texture homogeneity of GLCM	-0.013	-0.016	↓	0.011	0.003	↓
11	Texture entropy of GLCM	-0.005	0.010	↑	0.024	0.070	↑
12	Nuclei density	0	0	n/a	1	1	n/a

Table 4.7 Differences on the nuclear feature extracted from stained-based color-corrected image E without and with device-based color correction where the nuclear feature on reference image (WSI-A).

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B1A1$	$\Delta B2A1$	Diff.	$\Delta C1A1$	$\Delta C2A1$	Diff.
1	Area	2.583	2.560	↓	12.749	13.079	↑
2	Peripheral	0.420	0.383	↓	2.454	2.384	↓
3	Circularity	0.005	0.006	↑	0.001	0.002	↑
4	Long axis of elliptic fit	0.045	0.038	↓	0.437	0.379	↓
5	Short axis of elliptic fit	0.060	0.075	↑	0.314	0.349	↑
6	Ratio of short-long axis lengths	0.018	0.025	↑	0.008	0.011	↑
7	Contour complexity of cell	2.093	3.548	↑	0.591	0.467	↓
8	Texture ASM of GLCM	-0.008	-0.015	↓	-0.005	0.000	↑
9	Texture contrast of GLCM	-12.109	-9.975	↑	-33.156	-41.965	↓
10	Texture homogeneity of GLCM	-0.010	-0.028	↓	0.004	0.012	↑
11	Texture entropy of GLCM	0.047	0.099	↑	0.082	0.011	↓
12	Nuclei density	0	0	n/a	0	1	↑

4.4.4. Effect of combination of device- and stain-based color correction on intensity profiles

The intensity profiles on each image were evaluated by observing the intensities of nuclei and lymphocytes selected from each image. Two ROI images (64×64 pixels) containing a nucleus and a lymphocyte each were selected from image A and E. The details of these objects are described in Table 4.8. For each object, a set of ten intensity profiles were extracted from the original images (of A, B, and C), device-based color-corrected images (of B and C), stain-based color-corrected images (of A, B, and C), and image corrected by the combination of device- and stain-based color corrections (image B and C). In total, 40 images were gained in the process (10 subcases for each case). The details on each case can be seen in Table 4.9. Figure 4.22 and Figure 4.23 show luminance component of nuclei A-1 and E-1, along with the surrounding areas, in all its sub-cases. Here, the luminance value is based on the intensity on HIS color space, which calculated by averaging the R, G, and B values on each pixel. The luminance images show that the increment on contrast between the object and its surrounding is higher when stain-based color correction is performed.

Table 4.8 Objects used in the experiment.

No.	Case	Component	Image	Tissue Sample
1	A-1	Nucleus	A	Sample 1 (surgical sample)
2	A-2	Lymphocyte	A	Sample 1 (surgical sample)
3	E-1	Nucleus	E	Sample 2 (autopsy sample)
4	E-2	Lymphocyte	E	Sample 2 (autopsy sample)

Table 4.9 The sub-cases on each object.

No.	Sub-case ID	WSS	Color Correction
1	1	A	None
2	3	A	Stain-based
3	1	B	None
4	2	B	Device-based
5	3	B	Stain-based
6	4	B	Device-based + stain-based
7	1	C	None
8	2	C	Device-based
9	3	C	Stain-based
10	4	C	Device-based + stain-based

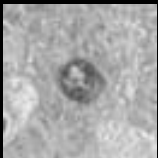
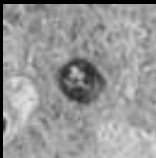
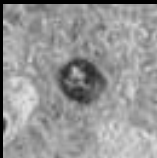
WSS	Sub-cases			
	w/o correction	Device-based color correction	Stain-based color correction	Device- and stain-based color correction
A		n/a		n/a
B				
C				

Figure 4.22 Intensity images of a nucleus from Case A-1 acquired by three different WSSs and the effect of color corrections on the images.

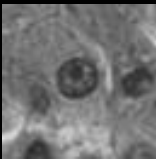
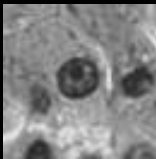
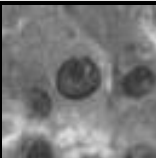
WSS	Sub-cases			
	w/o correction	Device-based color correction	Stain-based color correction	Device- and stain-based color correction
A		n/a		n/a
B				
C				

Figure 4.23 Intensity images of a nucleus from Case E-1 acquired by three different WSSs and the effect of color corrections on the images.

The effect of color correction on each image was analyzed by observing the intensity profile on a set of pixels in a line that cut through the nucleus or lymphocyte. The graphs in Figure 4.24 display the comparison of intensity profiles of the selected nuclei. For each figure, subfigures (a)–(f) compare the intensity profiles of non-color-corrected and color-corrected images captured by the same WSS, while subfigures (g)–(l) compare the intensity profiles of images captured by different WSSs. The details are described in Table 4.10 as follows.

Table 4.10 Description on sub-figures in Figure 4.24.

Subfigure	Case	Description
(a)	A-1	The difference on intensity profile between non-color corrected and stain-based color corrected image acquired by WSS-A.
(b)	E-1	
(c)	A-1	The difference on intensity profile between non-color corrected, device-based color correction, stain-based color correction, and combination of color correction on image acquired by WSS-B.
(d)	E-1	
(e)	A-1	The difference on intensity profile between non-color corrected, device-based color correction, stain-based color correction, and combination of color correction on image acquired by WSS-C.
(f)	E-1	
(g)	A-1	The intensity profile between original images acquired by WSS-A, WSS-B, and WSS-C.
(h)	E-1	
(i)	A-1	The intensity profile between stain-based color-corrected images acquired by WSS-A, WSS-B, and WSS-C.
(j)	E-1	
(k)	A-1	The intensity profile between subcase WSS-A (stain-based color corrected), WSS-B (device- and stain-based color corrected) and WSS-C (device- and stain-based color corrected).
(l)	E-1	

Overall, the color corrections increase the contrast between the nuclei or lymphocytes and the regions around the object. There are only slight differences on the intensity when combination of device- and stain-based color correction is implemented compared to a single stain-based color correction. This result confirm that the stain-based color correction can compensate the color differences due to the device variety.

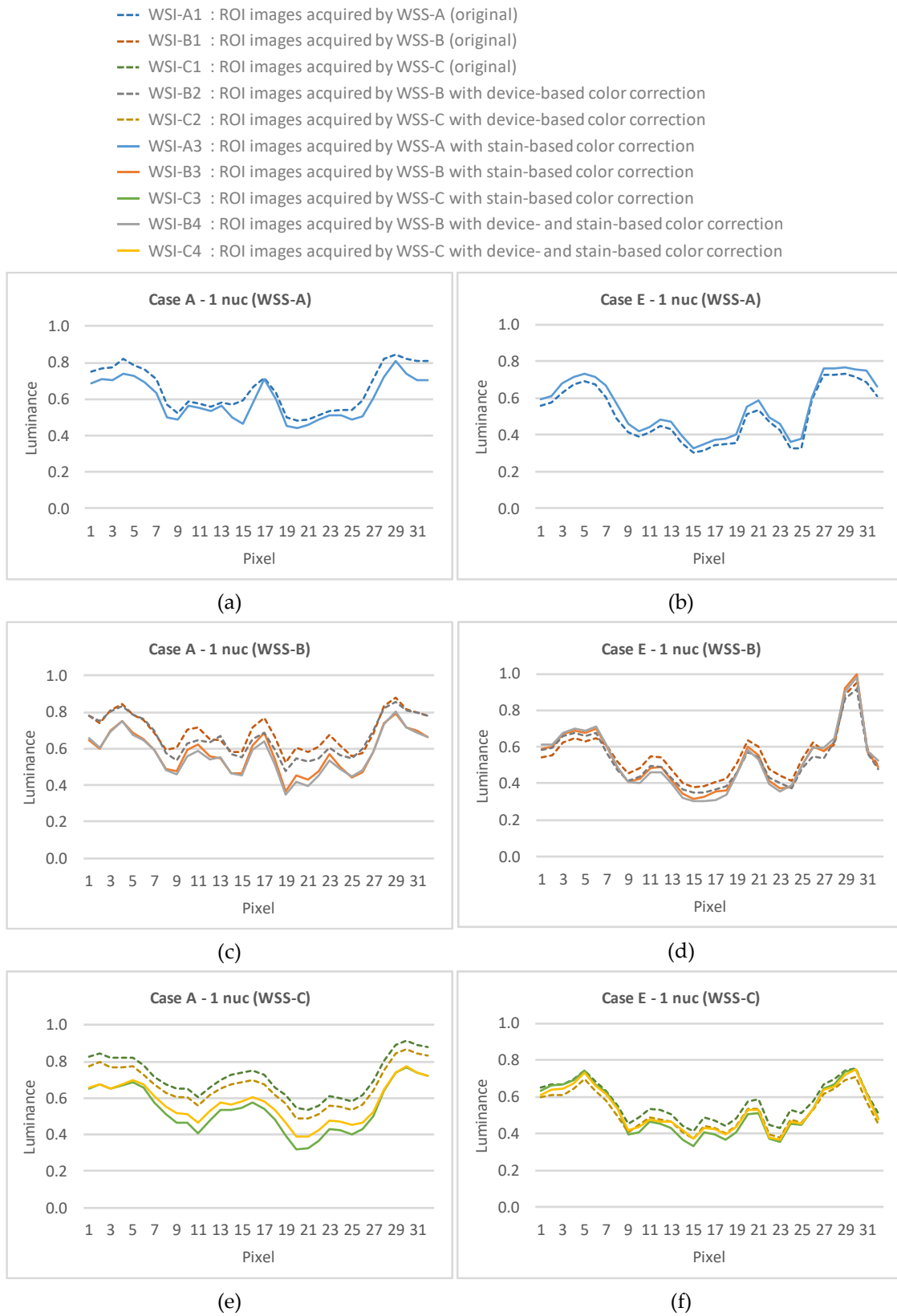


Figure 4.24 The comparison of intensity profiles of the selected nuclei.

- WSI-A1 : ROI images acquired by WSS-A (original)
- WSI-B1 : ROI images acquired by WSS-B (original)
- WSI-C1 : ROI images acquired by WSS-C (original)
- WSI-B2 : ROI images acquired by WSS-B with device-based color correction
- WSI-C2 : ROI images acquired by WSS-C with device-based color correction
- WSI-A3 : ROI images acquired by WSS-A with stain-based color correction
- WSI-B3 : ROI images acquired by WSS-B with stain-based color correction
- WSI-C3 : ROI images acquired by WSS-C with stain-based color correction
- WSI-B4 : ROI images acquired by WSS-B with device- and stain-based color correction
- WSI-C4 : ROI images acquired by WSS-C with device- and stain-based color correction

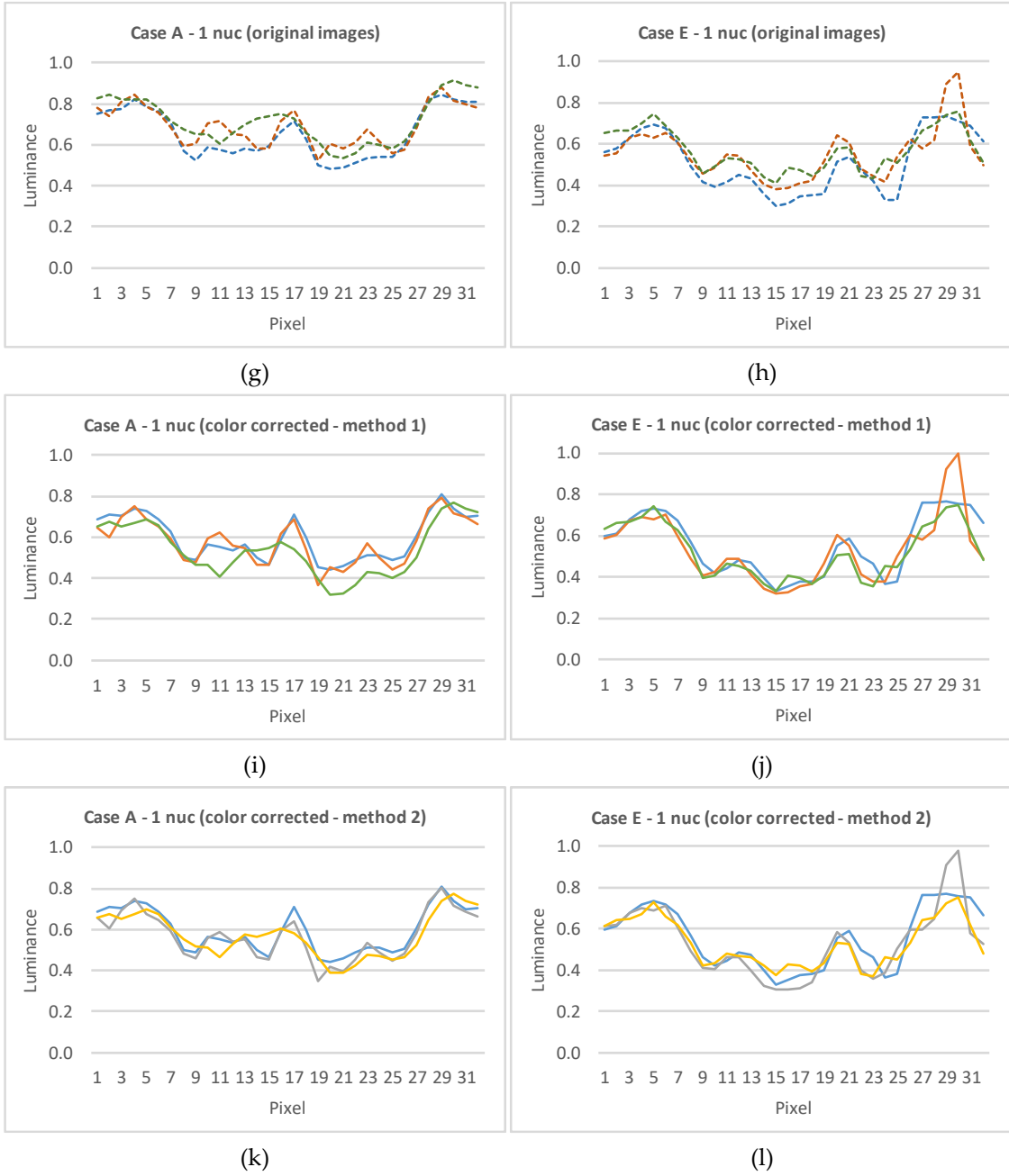


Figure 4.24. (continued).

4.5. Discussions

Color correction on WSI is essential to reduce the color difference on tissue samples due to staining process. In this chapter, the effectiveness of several color correction method on nuclei quantification from WSIs captured by different WSSs were analyzed. In the case of stain-based color correction, the results show that the method is more effective on nuclei's morphological features but can be less effective on texture-related features. It has been confirmed that the features related to the shape of nuclei, such as nuclear area, peripheral, circularity, and ratio of short & long axis lengths of elliptic fit are stable for color correction, whereas textural features tend to be affected by the device difference. Applying the color correction on histopathological image analysis software improves the reliability of the quantification process of the tissue features. Color correction increases the image's contrast, which allows better detection on objects having high contrast from its surrounding areas, such as nuclei and lymphocytes. Nevertheless, variations of the quantified features still exist, especially on some particular type of features due to its sensitiveness on image contrast. For this, further investigations on the image quality, such as on image sharpness and device-based color correction, are necessary. In addition, the focus variations in the scanning process cannot be ignored, and image quality is affected. Digital correction techniques on image sharpness, such as modulation transfer function (MTF) based correction or blind deconvolution, may be implemented to overcome existing differences.

In the case of device-based color correction, the results indicate that the device-based color correction is less effective compared to the stain-based color correction. The effect of device-based color correction on the color chart slide image and the histopathological image is also vary. One possible cause of this result is due to differences on spectra between the color on the color chart slide and the color from the stain in the tissue. Moreover, the materials on the color chart slide have different density compared to the tissue specimens. These phenomena were also stated in [56]. Combining the device- and stain-based color correction does not produce high impact compared to a single stain-based color correction. Nevertheless, device-based color correction by using color calibration slide that is able to mimic tissue characteristics, as introduced by Clarke et al. [56], can be an interesting topic in the future.

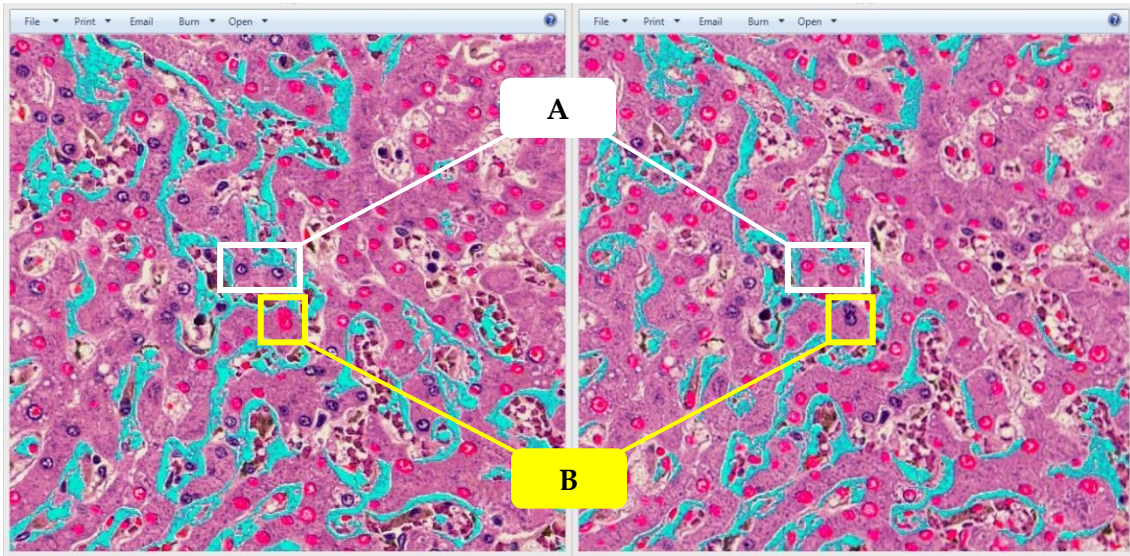
Chapter 5

Sharpness correction and its effect on quantitative pathology

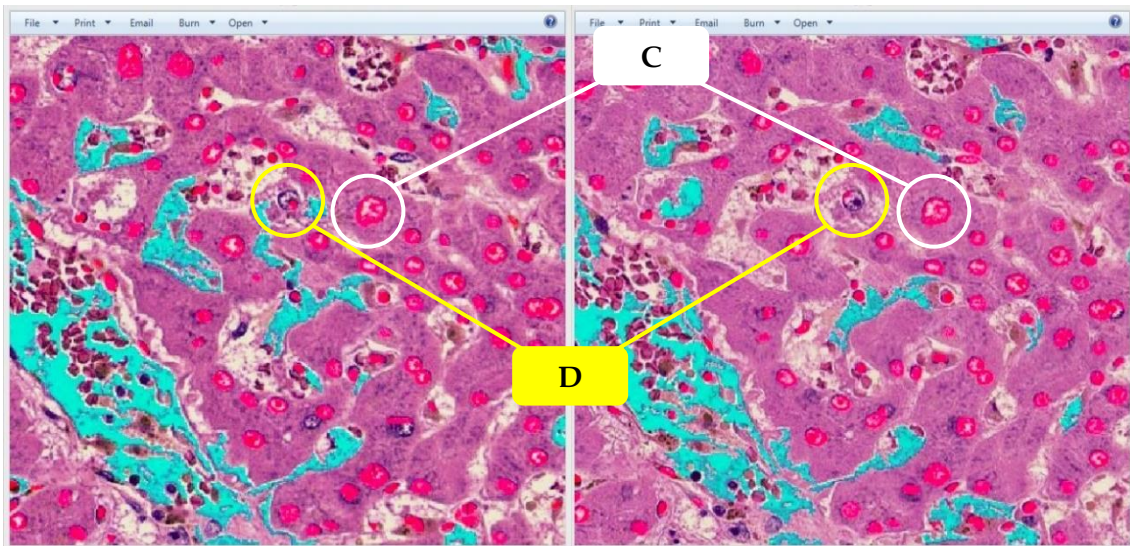
One of the main objective of CAD system development is to extract quantitative measurements from the tissue components and provide the information to the pathologists. A reliable system has to have ability in delivering similar results even though the specimens are digitized by using different scanner devices. Nevertheless, variety of WSSs will lead to difference on image quality. This chapter focuses in the sharpness quality of histopathological images.

5.1. Sharpness on histopathological images acquired by different WSSs

Image sharpness is one of the most essential factors determining the quality of an image. This is also applied in histopathological images. It is difficult to make observation on an out-of-focus image manually. The influence of image sharpness becomes more essential when the histopathological image is analyzed through image analysis implementation. Figure 5.1 illustrates how image quality, especially image sharpness, plays important part in image analysis. The figure shows two regions in a liver tissue where the images were digitized by different WSSs (WSI-A and WSI-C), where some tissue regions detected as nuclei and fiber were overlaid with magenta and cyan colors, respectively. Even though the image's quality looks similar in naked eye, the result of object detections is varied. The case A and B in Figure 5.1 (a) show nuclei that are detected in one image but are not detected by the other image. In the meantime, Case D in Figure 5.1 (b) shows nucleus that is detected partially on both images. Such difference need to be overcome in order to produce a reliable CAD system.



(a)



(b)

Figure 5.1 Effect of image quality on quantitative pathology. (a) In case A, the two nuclei are only detected on ROI image from WSI-A (right) while in case B, the nucleus is only detected on ROI image from WSI -C (left). (b) Both in case E and F, the nuclei are detected on both ROI images. However, the masking area in case F are different between the two images. Image on the left side is from WSI-C while image on the right side is from WSI-A.

5.2. MTF-based correction on histopathological images

5.2.1. Introduction to Wiener filter

Image restoration is a process to restore or compensate image that have been degraded by a degradation function. The image degradation and restoration can be modeled as described in Figure 5.2 below.

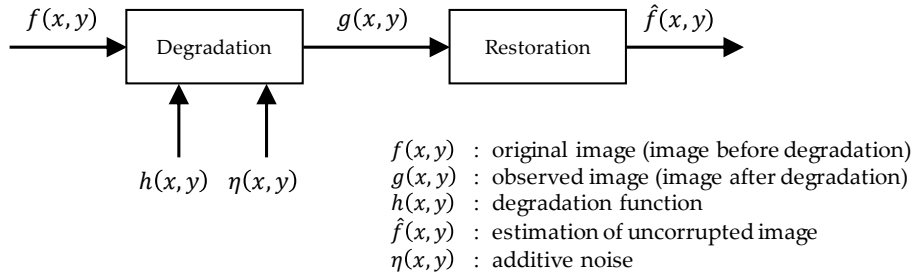


Figure 5.2 Image degradation and restoration model.

Mathematically, the observed image can be defined in the spatial domain by

$$g(x, y) = h(x, y) * f(x, y) + \eta(x, y) \quad (5.1)$$

or in frequency domain as

$$G(u, v) = H(u, v)F(u, v) + N(u, v) \quad (5.2)$$

There are various types of image restoration filters to restore degraded image. Wiener filter is one of the restoration methods. The Wiener filter considers images and noise as random variables and assumes both are uncorrelated. The method also assumes that the intensity levels in the estimated image are a linear function of the levels in the degraded image. Wiener filter then estimate the restored image by minimizing the mean square error with the uncorrupted image, which is why the method is commonly referred as minimum mean square error or least square error filter. The minimum error is measured by

$$e^2 = E \{ (f - \hat{f})^2 \} \quad (5.3)$$

where $E\{\cdot\}$ is the expected value of the argument. Based on the considered conditions mentioned earlier, the estimation of the uncorrupted image can be expressed in the frequency domain as

$$\hat{F}(u, v) = W(u, v)G(u, v) \quad (5.3)$$

where $W(u, v)$ is the Wiener filter which is defined as

$$W(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + \frac{S_\eta(u, v)}{S_f(u, v)}} \quad (5.4)$$

Here, $H(u, v)$ is the degradation function, $H^*(u, v)$ is the complex conjugate of $H(u, v)$, $|H(u, v)|^2$ is equal to $H^*(u, v) H(u, v)$, $S_\eta(u, v)$ is the power spectrum of the noise and equal to $|N(u, v)|^2$, and $S_f(u, v)$ is the power spectrum of the undegraded image and equal to $|F(u, v)|^2$. However, the power spectrum of the undegraded image is seldom known. In the meantime, the power spectrum of the noise is constant when dealing with spectrally white noise. Therefore, an approach is often utilized to deal with these unknown quantities, which is by specifying constant value K so that the equation (5.4) becomes

$$W(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + K} \quad (5.5)$$

Note that when there is no noise, K equals to 0 and the filter becomes inverse filter [54].

5.2.2. Histopathological-images-derived MTF filter

Image sharpness is highly related to the depth of field during object acquisition. One way to correct sharpness on image is by using modulation transfer function (MTF)-based correction. In this work, a filter is estimated and derived from a pair of histopathological images of the same object, where both images were obtained by the reference WSS and the target WSS. The workflow on deriving the filter are described in Figure 5.3.

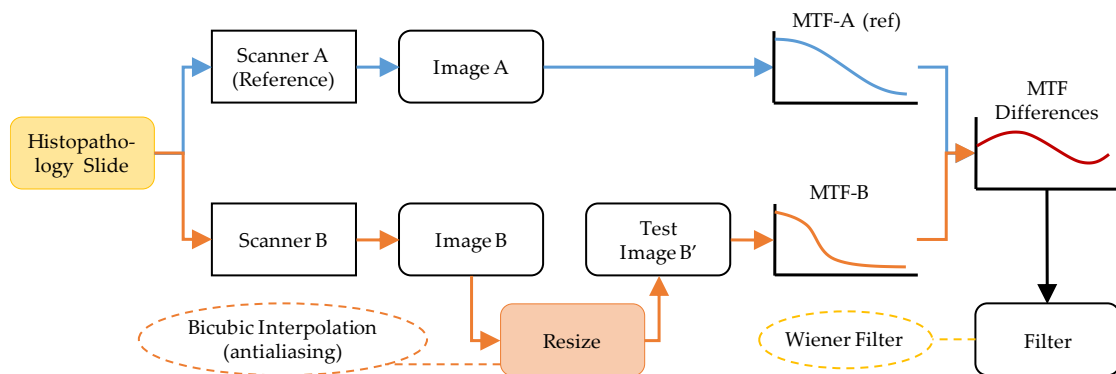


Figure 5.3 Workflow to derive MTF-based correction on histopathological images, where the filter is based on a reference histopathological image.

Figure 5.4 shows the model of the process. Here, WSS₁ is the reference scanner and it is assumed that the produced image $g_1(x, y)$ does not contain noise as it is the reference image. The focus is to estimate $\hat{g}_1(x, y)$ based on the MTF difference $M(u, v)$ between MTF of $g_1(x, y)$ and $g_2(x, y)$, which can be defined in frequency domain as

$$\hat{G}_1(u, v) = M(u, v)G_2(u, v) \quad (5.6)$$

$$M(u, v) = \frac{H_1(u, v)}{H_2(u, v) + \sigma} \quad (5.7)$$

where $H_1(u, v)$ and $H_2(u, v)$ are the MTF of WSS₁ and WSS₂ respectively while σ is a small scalar value to reduce amplification effect in higher frequency.

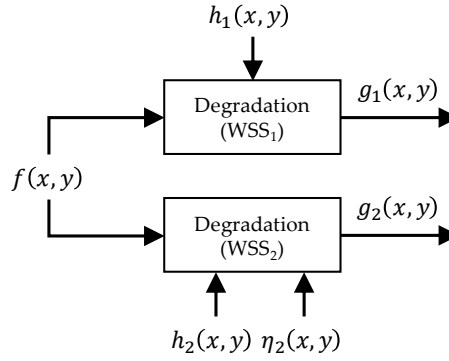


Figure 5.4 Workflow to derive MTF-based correction on histopathological images, where the filter is based on a reference histopathological image.

As stated earlier, the MTFs of both imaging systems are derived from the histopathological images. The process started by obtaining WSIs of a tissue specimen by the reference and target WSSs and confirming the size of the target image is similar to the reference image. After that, the appropriate region on the WSIs was selected to derive the filter. Once the area covered in the images were fixed, the MTFs of the two images is obtained by deriving the magnitude of Fourier transform on each image and calculating the average of the magnitude having the same distance to the center point on each direction, as shown in Figure 5.5. Here, S equals to the half of the image size. The average value for each point having the same distance s to the center point is calculated for each angle θ having range of $[0, 359]$. To smoothen the curve, polynomial fitting is implemented. Calculating the difference between the two MTFs of the images will produce the 1-dimension of the filter. The 2-dimensional filter is produced by rotating

the 1D filter having the highest frequency as the center point. The region having distance farther S is set to have the same value with the average amplitude at $s = S$.

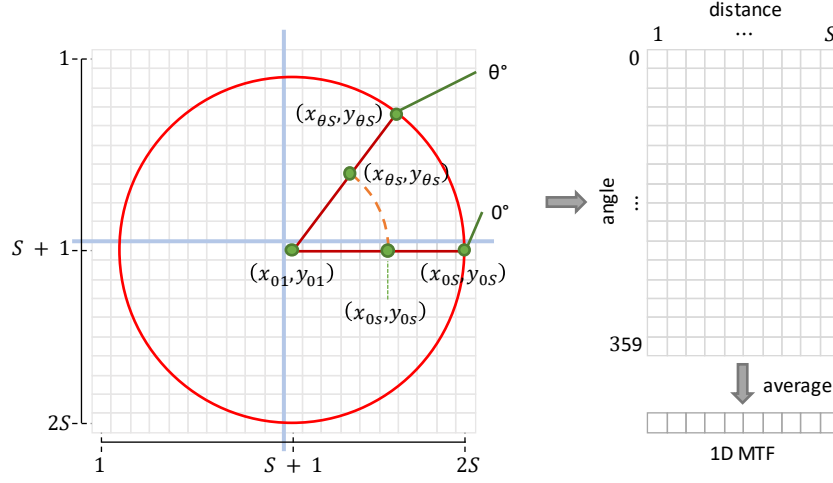


Figure 5.5 Illustration on the process of averaging the magnitude of the image's Fourier transform for each angle in order to generate 1D MTF.

5.2.3. Evaluation on MTF-based correction on histopathological images

To evaluate the effectiveness of this method, the Wiener filter derived from the above process was implemented on several images captured the target WSSs and nuclear detection module was applied on the images. The results were compared to the number of detected nuclei on the reference images. Figure 5.6 below shows ten histopathological images used for the study. The size of each image is 1024×1024 pixel. Images A—D are from surgical resected liver specimen, images E—G are from an autopsy liver specimen, and images H—J are from another autopsy liver specimen. Each specimen was digitized by using three different WSSs, WSS-A, WSS-B, and WSS-C, having images acquired by WSS-A as the reference image. The MTF-based filter was derived from image A. Figure 5.7 show the 2D MTF of each WSS in the first row, the 2D MTF-based filters in the second row, and the 1D MTF differences between WSS-B and WSS-C with WSS-A in the third row. The derived MTF-based filter was implemented on images B—J. The number of detected nuclei on the images, before and after MTF-based correction are compared to find out the effect of the correction.

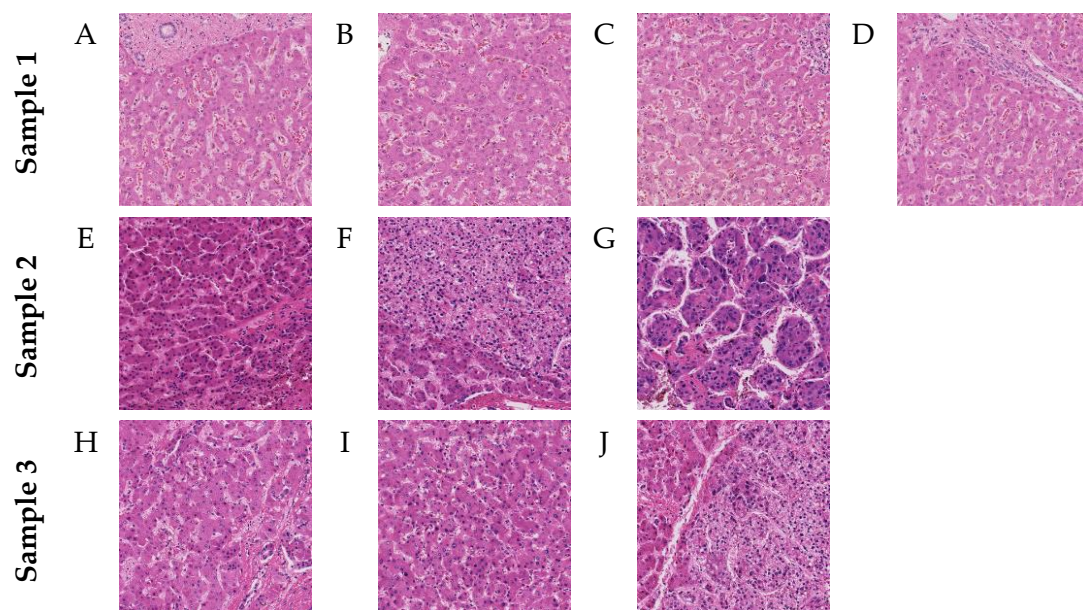


Figure 5.6 ROI images used for the experiment.

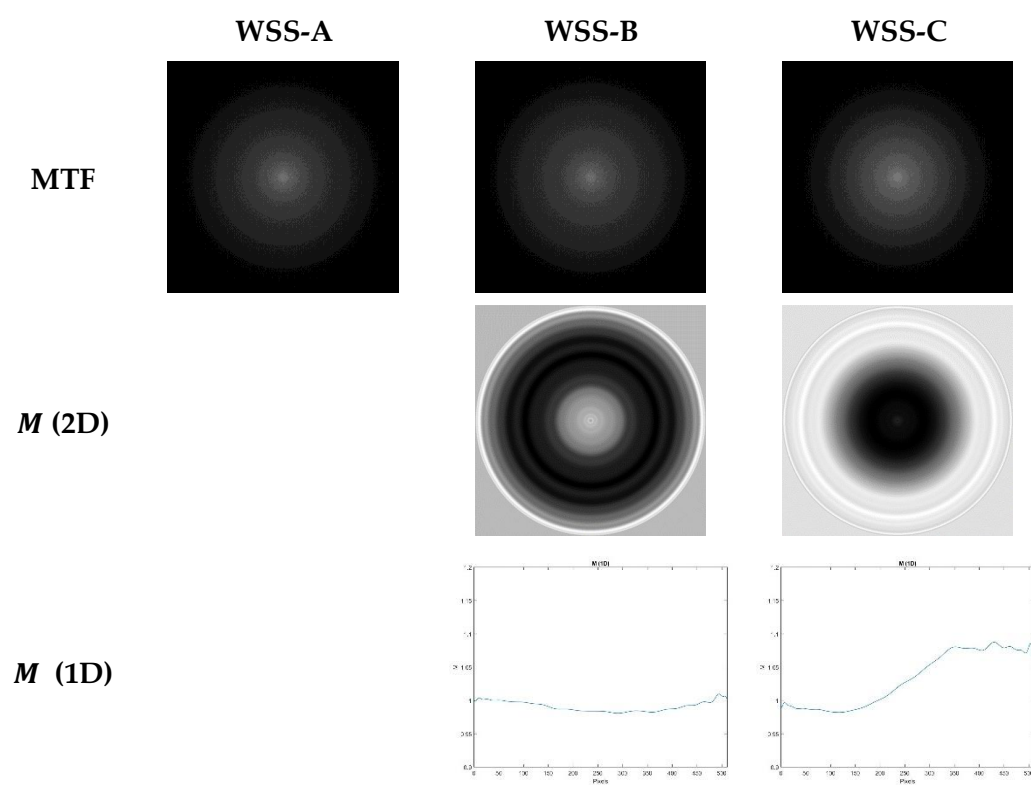


Figure 5.7 The 2D MTF of each WSS, the 2D MTF-based filters, and the 1D MTF differences between WSS-B and WSS-C with WSS-A.

Figure 5.8 shows several ROI images after MTF-based correction. From visual observation, it is difficult to see the difference between the images before and after MTF-based correction. Figure 5.9 shows the comparison between the number of detected nuclei. Implementing the MTF-based correction on histopathological images does not reduce the difference on the quantitative pathology. One of the reason is the MTF filter is dependent on the histological images employed to derived filter. Furthermore, the autofocus function on WSS may cause WSIs of a tissue captured by the same WSS at different time produces image having different focus as shown in Figure 5.10. The orange marks show some objects having different shape. The white area in Figure 5.10 (b) is wider compared to the same area in Figure 5.10(a), while some red blood cells that can be seen in Figure 5.10(d) look like artifacts in Figure 5.10(c).

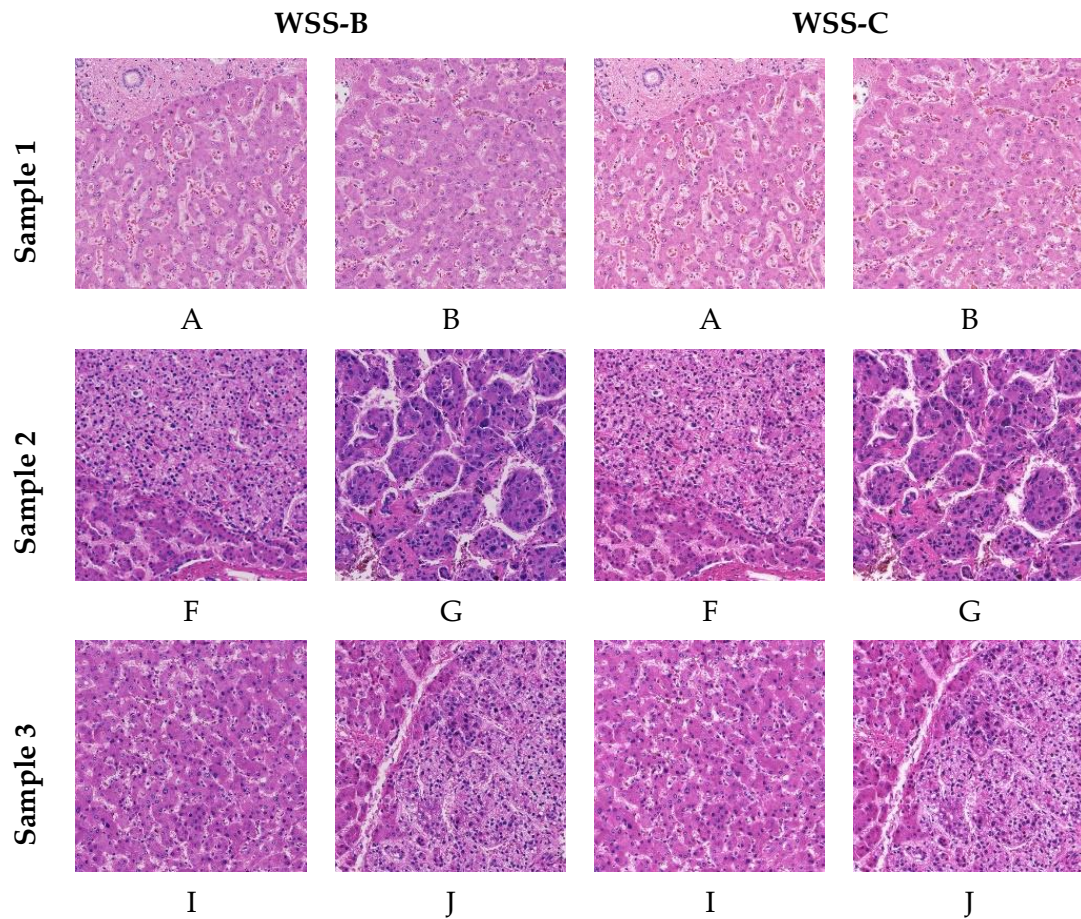


Figure 5.8 Several ROI images after MTF -based correction.

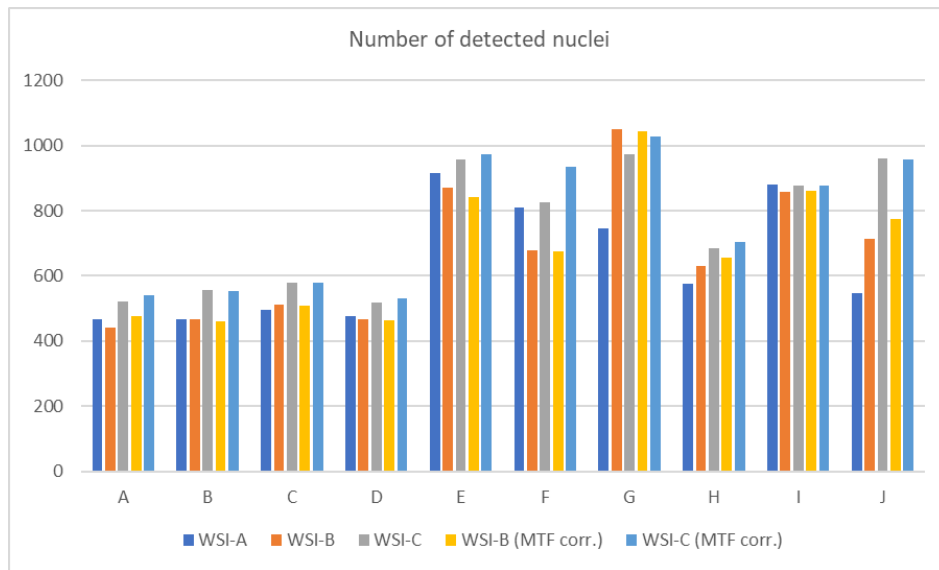


Figure 5.9 Comparison on the number of nuclei detected on WSIs before and after MTF-based correction.

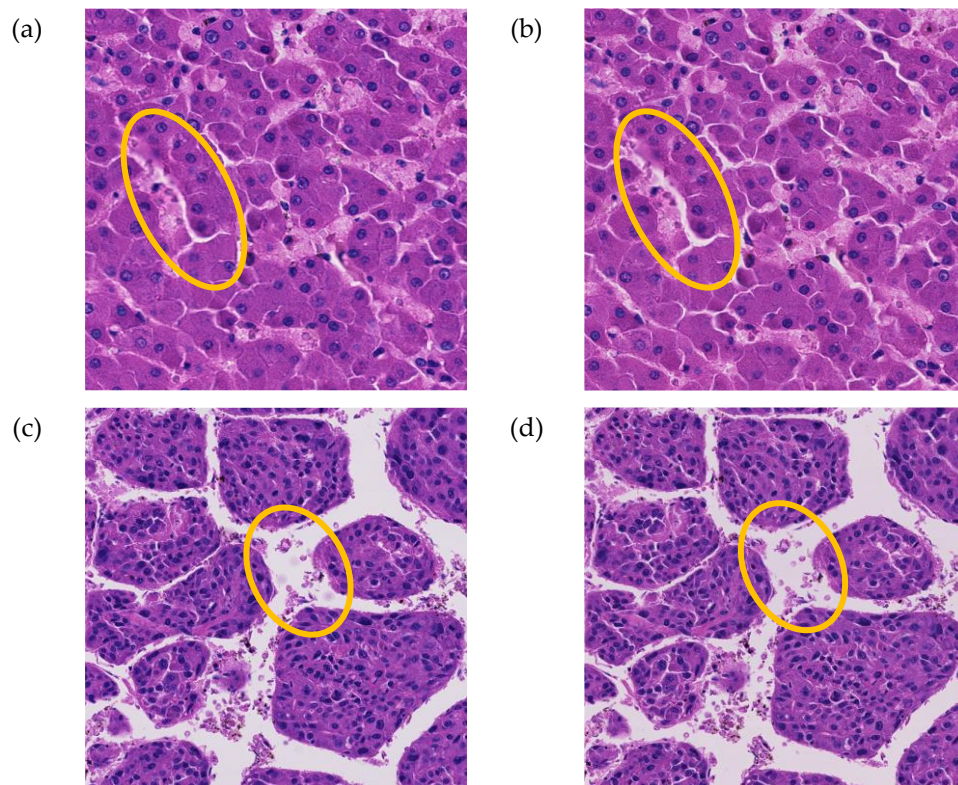


Figure 5.10 WSIs of a tissue scanned by a WSS at different time, showing different shape of objects due to difference depth of field during scanning process.

5.3. Gaussian model sharpness correction on histopathological images acquired by different WSSs

5.3.1 Sharpness correction based on Gaussian model point spread function

The previous subsection has shown that differences on the depth of field during digitizing the tissue specimen can cause variety on image sharpness or even focus blur. Furthermore, the sharpness variety may also be taken place on the images acquired by the reference scanner. Therefore, it becomes more appropriate to carry out image restoration, which in this case deblurring process, based on the sharpness level of the histopathological images instead of the device-based (either reference or non-reference device).

One way to overcome the optical blur problem mentioned above in the histopathological images is by performing deconvolution based on Gaussian model point spread function (PSF). Deconvoluting image based on the Gaussian model PSF requires optimization on the size of Gaussian (σ), which will be determined based on the quality for each histopathological image. This section focuses on the investigation of the implementation of sharpness correction on WSIs captured by different devices and its effect on the histopathological images and quantified features.

5.3.2. Experiment workflow

The experiment basically the extended version of experiment introduced in Subsection 4.2. Two sets of two ROI images were prepared for this experiment, where the two images were utilized in the Subsection 4.2, which are images G and K. Both images are diagnosed as HCC cases. Sharpness correction was applied on one of the two ROI image sets. The sharpness correction was performed by performing deconvolution on the image, where the target images were deconvolved with a set of Gaussian model PSFs having different sizes (σ). Figure 5.11 shows an example of the implementation of Gaussian model PSFs having several σ values on images G.

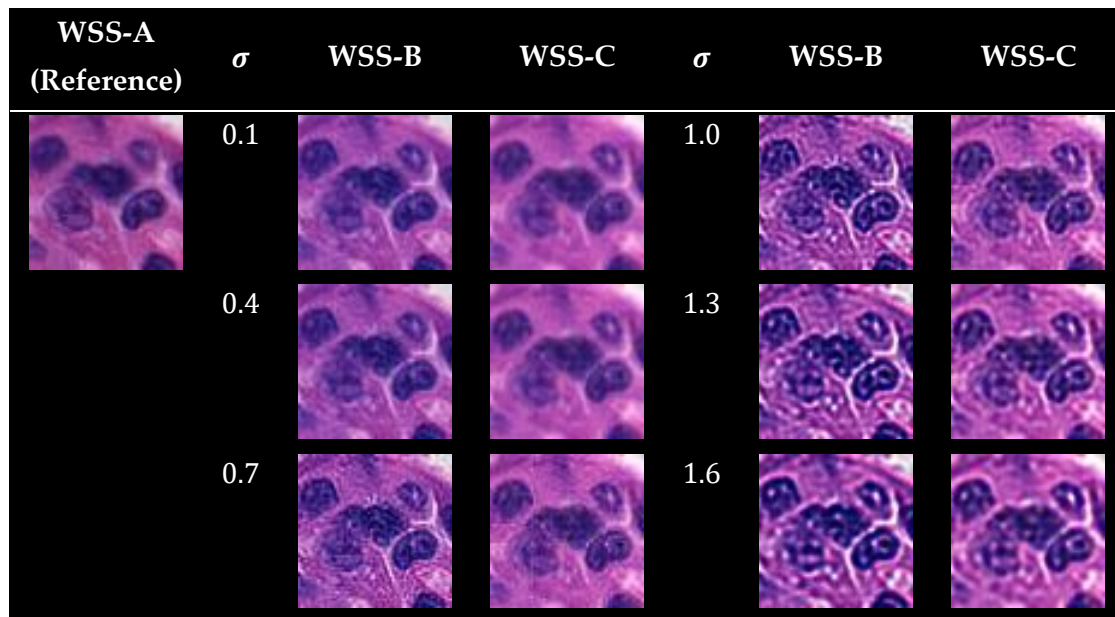


Figure 5.11 The results of image deblurring with Gaussian model PSFs having several σ values.

The deblurred images were observed manually and deblurred image having the most similar quality with the reference image is chosen as the corrected image. Stain-based color correction and feature quantification were then carried out on both sets. The results on nuclei detection, quantified features and image's intensity profiles were then compared between the image without and with sharpness correction in order to evaluate the effect of sharpness correction on histopathological images captured by different slide scanners. The overall workflow is shown in Figure 5.12.

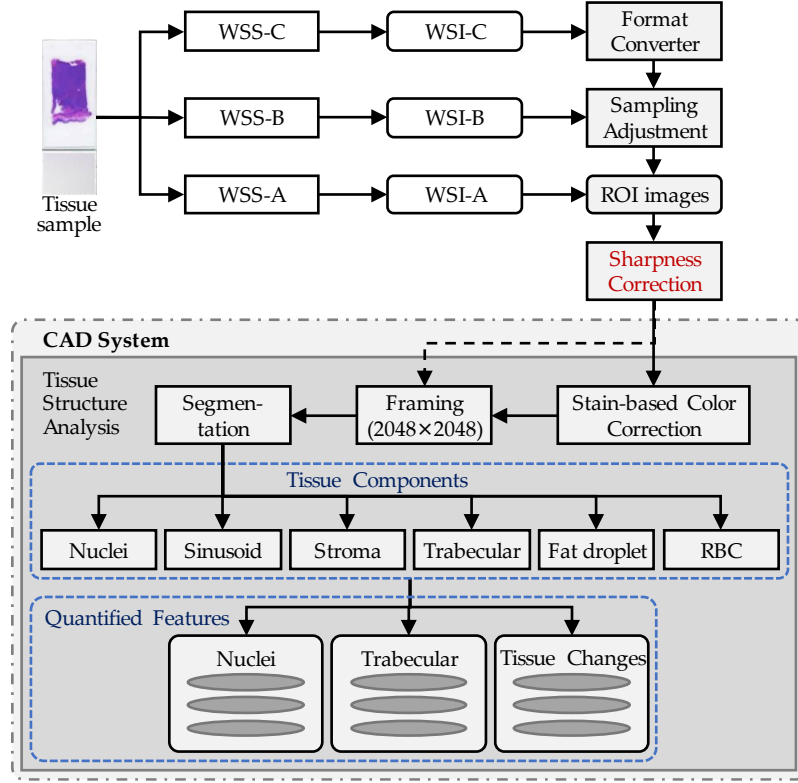


Figure 5.12 Workflow of experiment on sharpness correction. Two ROI images from two liver specimens were used, where each specimen was scanned by WSS-A, WSS-B, and WSS-C. Sharpness correction is implemented ahead of the stain-based color corrections on the ROI images.

5.3.3. Effect of sharpness correction on nuclei detection

The effect of sharpness correction on the nuclei detection is shown in Figure 5.13. Between the four cases (two corrected images scanned by two target WSSs), the difference on the number of nuclei detected only reduced in the case of image G, scanned by WSS-B. Nevertheless, the changes on the number of nuclei detected are relatively small in all cases, as shown in Table 5.1. Here, $\Delta B2A2$ and $\Delta C2A2$ means the result difference between non-corrected images captured by WSS-B and WSS-C, respectively, with the images captured by WSS-A. Meanwhile, $\Delta B3A2$ and $\Delta C3A2$ means the result difference between sharpness-corrected images captured by WSS-B and WSS-C, respectively, with the images captured by WSS-A. Note that stain-based color corrected

was implemented in all cases in this experiment. Figure 5.14 and Figure 5.15 shows the effect of sharpness correction on each case.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

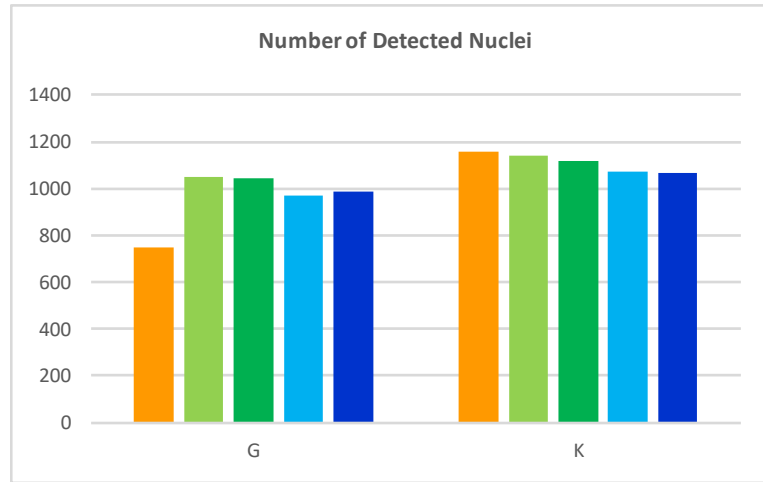


Figure 5.13 Comparison on the number of nuclei detected on image before and after sharpness correction.

Table 5.1 Differences on the number of nuclei detected between images before and after sharpness correction.

Image	WSI-B			WSI-C		
	$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
G	303	298	↓	226	243	↑
K	15	41	↑	85	91	↑

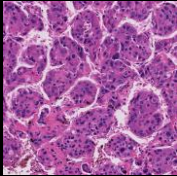
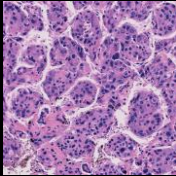
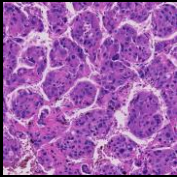
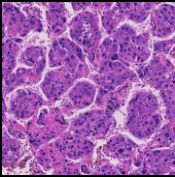
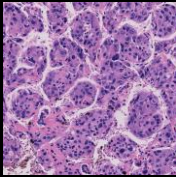
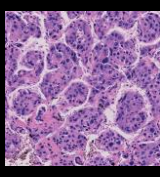
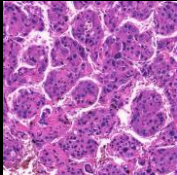
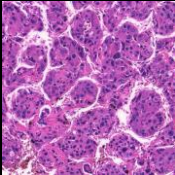
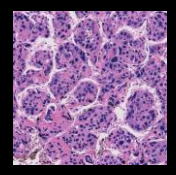
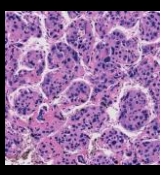
WSS	w/o stain-based color correction		with stain-based color correction	
	w/o sharpness correction	with sharpness correction	w/o sharpness correction	with sharpness correction
A		n/a		n/a
B				
C				

Figure 5.14 Image correction on image G.

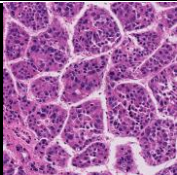
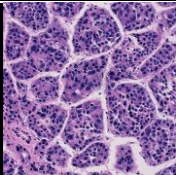
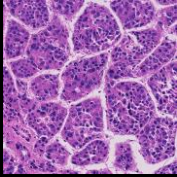
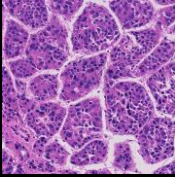
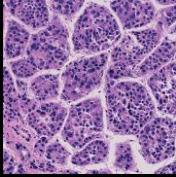
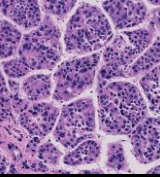
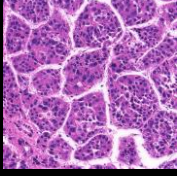
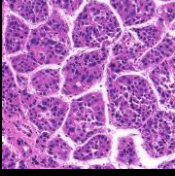
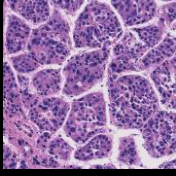
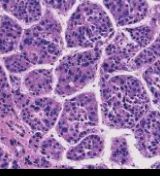
WSS	w/o stain-based color correction		with stain-based color correction	
	w/o sharpness correction	with sharpness correction	w/o sharpness correction	with sharpness correction
A		n/a		n/a
B				
C				

Figure 5.15 Image correction on image K.

5.3.4. Effect of sharpness correction on intensity profiles

To evaluate the effect of sharpness correction on intensity profiles, four nuclei from two ROI images (images G and K) were selected. The nuclei from image G are labeled as nuclei B and C, while the nuclei from image K are labeled as nuclei E and F. Figure 5.16 shows the nuclei in gray-level and RGB images, where the nuclei are masked with green color in the RGB images.

The comparison on intensity profiles of nuclei B, C, E, and F are shown in Figure 5.17, Figure 5.18, Figure 5.19, and Figure 5.20 respectively. Subfigures (a) compare the intensity profiles of images without sharpness correction and subfigures (b) compare the intensity profiles of images with sharpness correction. Meanwhile, subfigures (c) and (d) compare the intensity profiles without and with sharpness correction on image acquired by WSS-B and WSS-C, respectively. Here, images acquired by WSS-A were used as the reference images. In all cases, the intensity profiles before and after sharpness correction are relatively similar, having range of differences on intensity values between $[0, 0.0170]$.

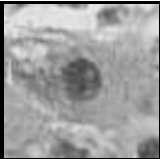
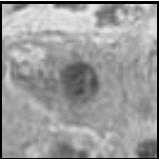
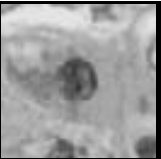
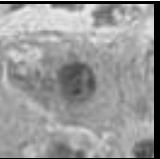
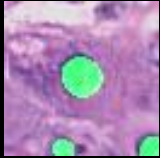
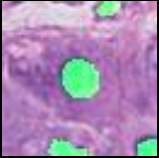
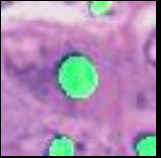
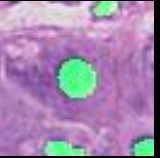
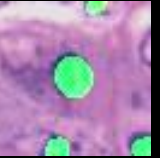
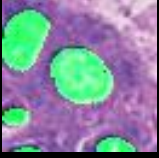
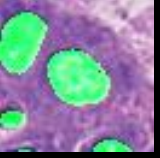
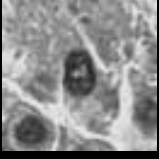
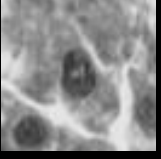
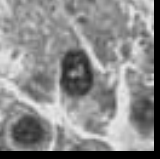
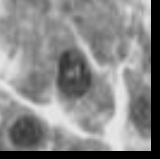
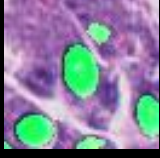
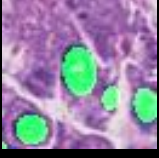
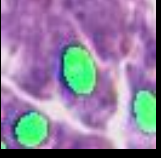
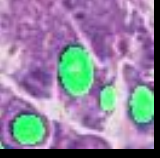
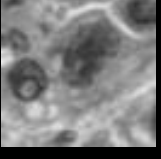
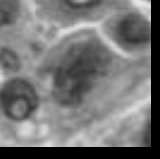
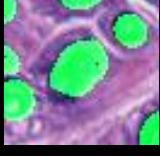
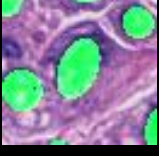
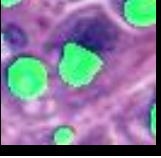
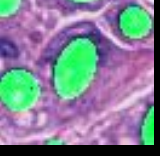
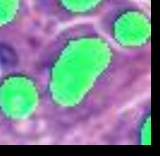
No.	Img.	Nuc.	NHT	w/o sharpness correction		with sharpness correction	
				NXR	3DH	NXR	3DH
(1)	G	B					
(2)							
(3)		C					
(4)							
(5)	K	E					
(6)							
(7)		F					
(8)							

Figure 5.16 The WSIs of three tissue specimens used for the experiment. Each specimen was digitized thrice by three different WSSs.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- - WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- - WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

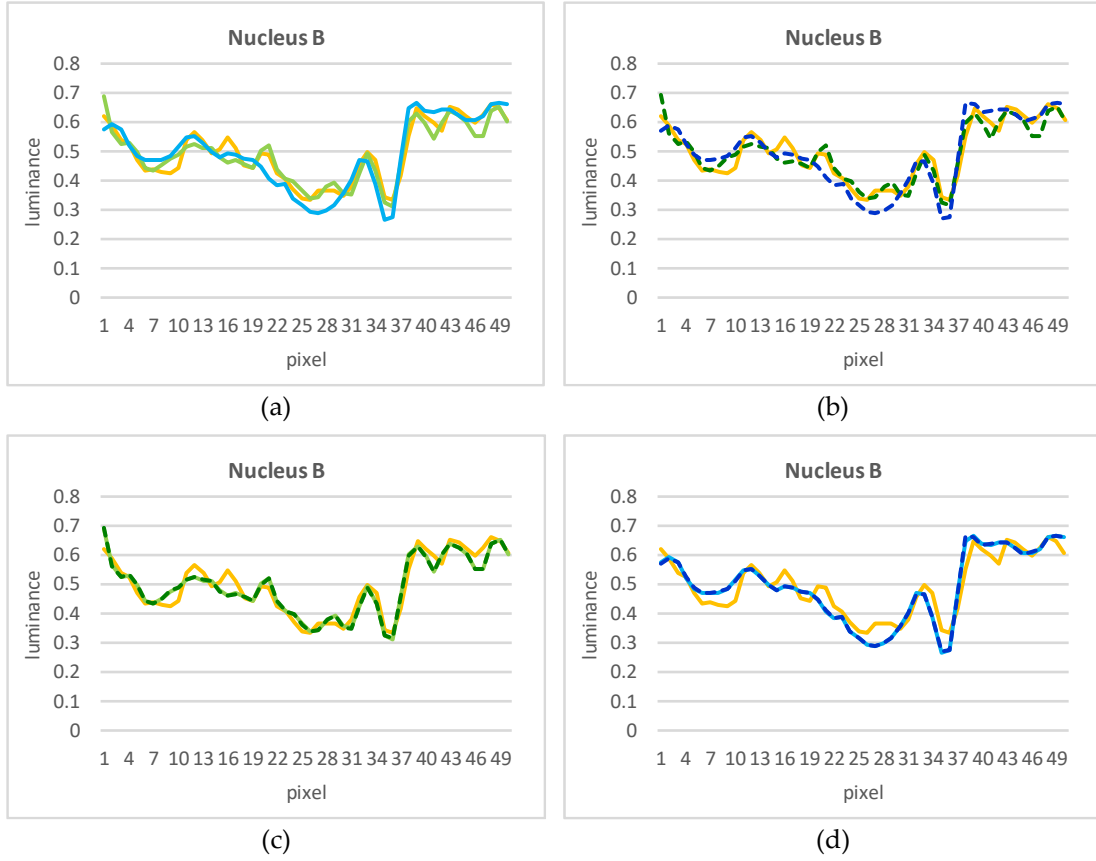


Figure 5.17 Comparison on the intensity values of nucleus B. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are $[0, 0.0039]$ and $[0, 0.0170]$, respectively.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- - WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- - WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

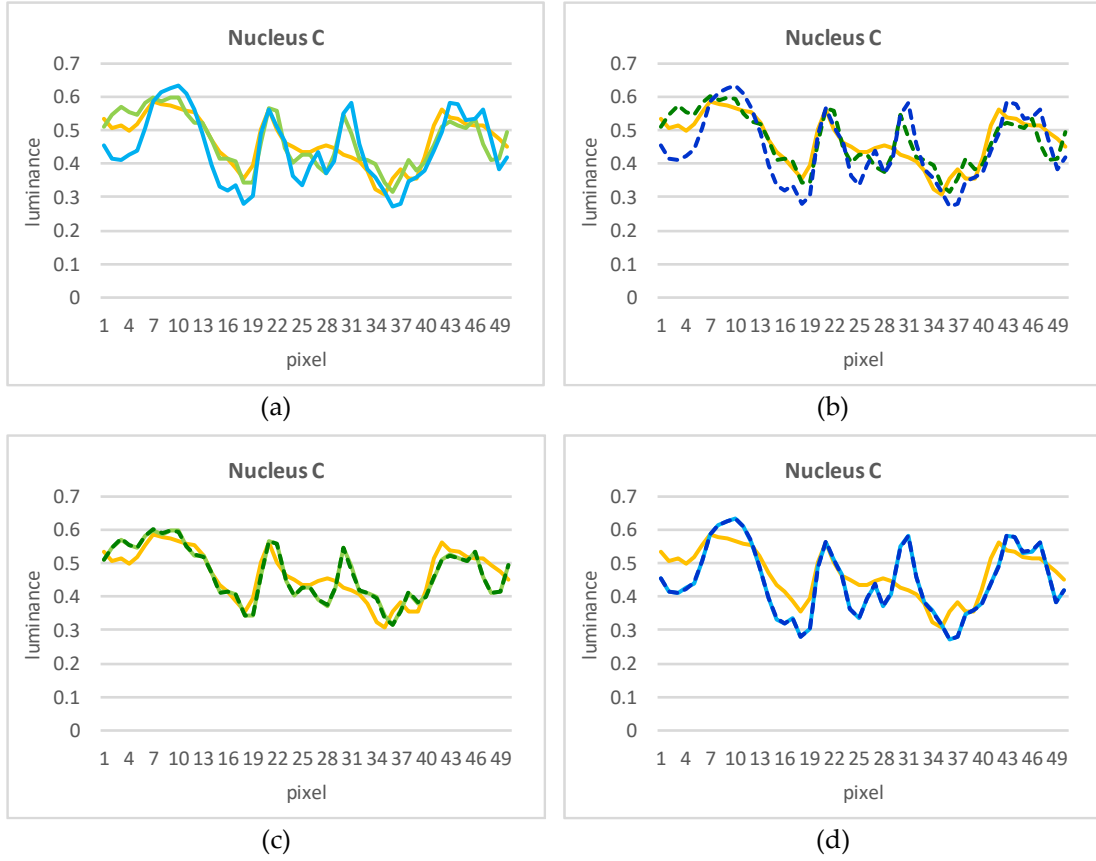


Figure 5.18 Comparison on the intensity values of nucleus C. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are $[0, 0.0118]$ and $[0, 0.0065]$, respectively.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- - WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- - WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

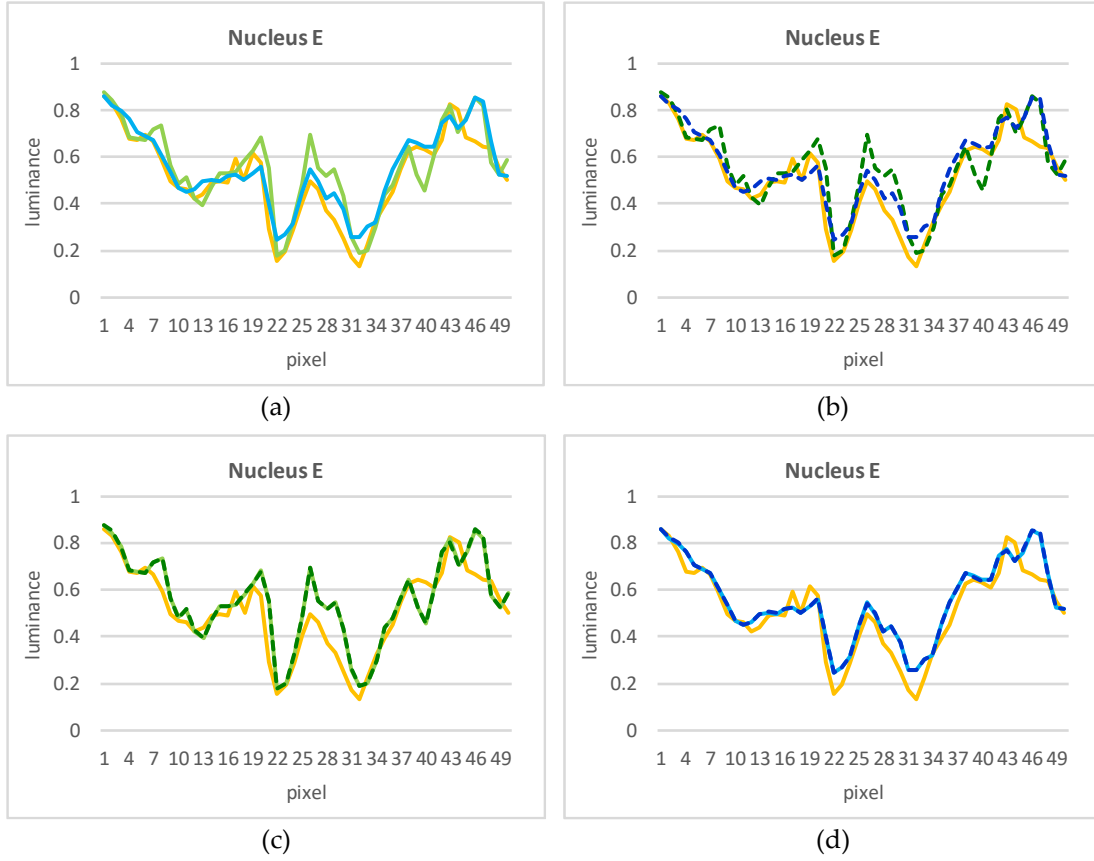


Figure 5.19 Comparison on the intensity values of nucleus E. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are $[0, 0.0157]$ and $[0, 0.0105]$, respectively.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- - WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- - WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

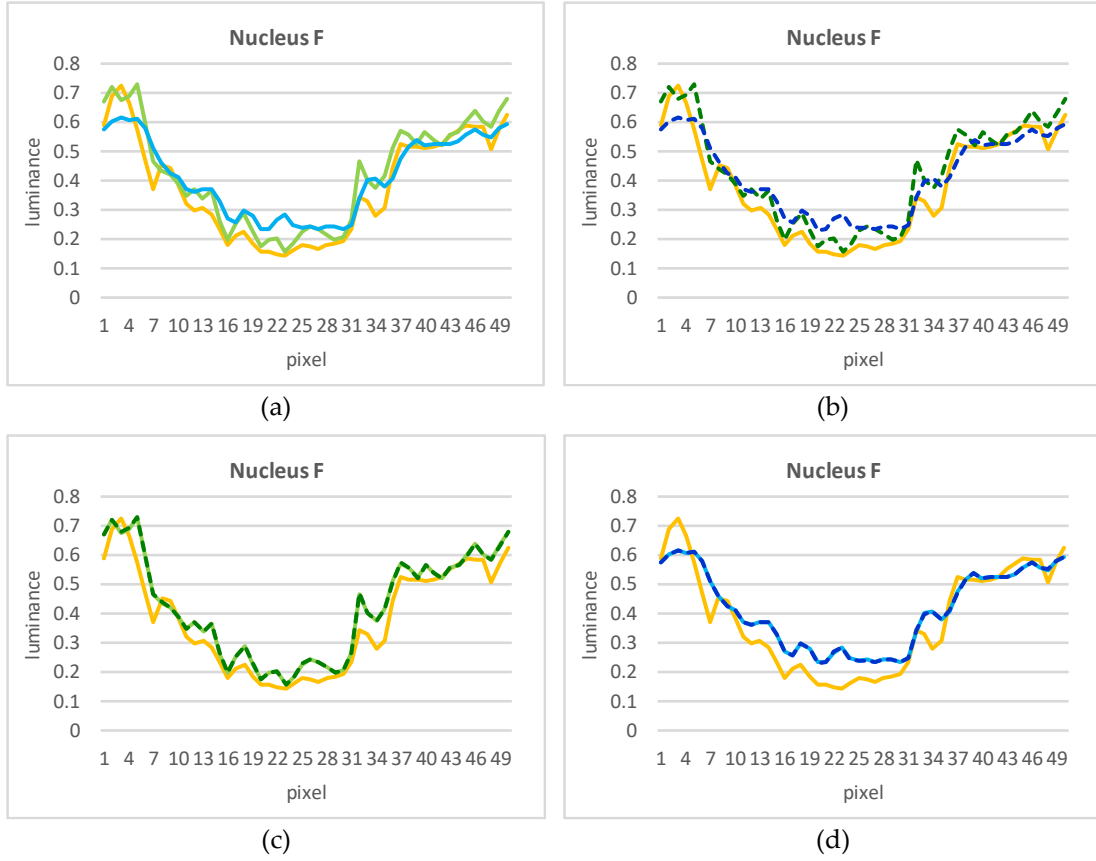


Figure 5.20 Comparison on the intensity values of nucleus F. The ranges of difference on intensity profiles from images captured by WSS-B and WSS-C before and after sharpness correction are $[0, 0.0052]$ and $[0, 0.0039]$, respectively.

5.3.5. Effect of sharpness correction on quantified features

To further evaluate the four nuclei mentioned in the previous subsection, the nuclei quantification for each single nucleus is analyzed. The results are shown in Figure 5.21. The graphs in (a), (c), (e), (g), (i), (k), (m), (o), (q), (s), and (u) are the comparison between images acquired by WSS-B (without and with sharpness correction) with the reference images. While the other graphs are the comparison between images acquired by WSS-C. In total, 11 features are extracted from a single nucleus, consisting the nuclear area, perimeter, circularity, long and short axes, ellipticity, contour complexity, and four features related to GLCM (ASM, contrast, homogeneity, and entropy). There are different effects of sharpness correction on nuclear quantification. There are less effects found in nuclei B and C in most features, especially in morphological related features. More positive effects can be found on nucleus F, especially on image captured by WSS-C. Without the sharpness correction, the nucleus was only detected partially on the image. By correcting the sharpness, the nucleus can be detected better. The details of differences on the nuclei quantification on each case are shown in Table 5.2–Table 5.5.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

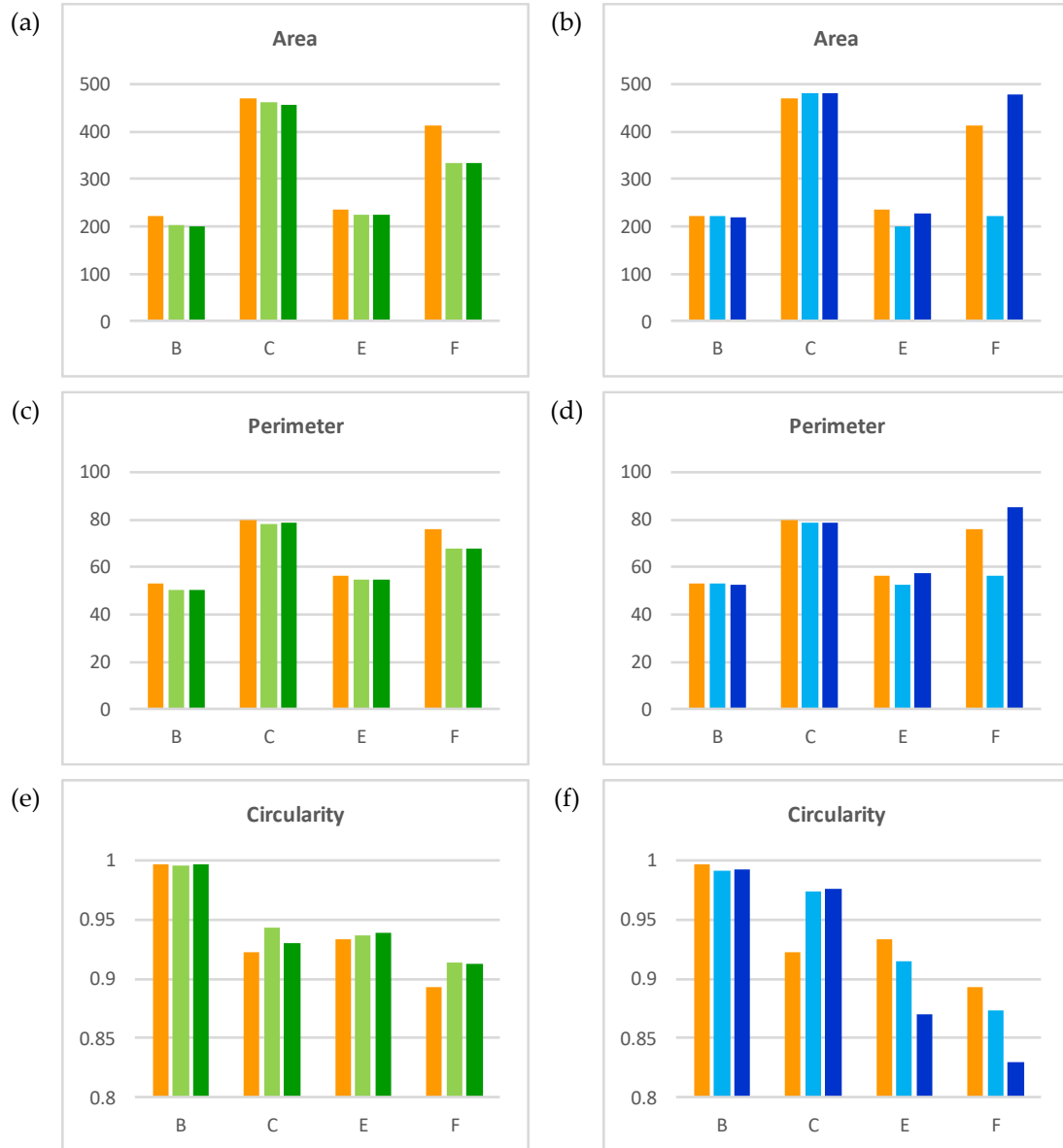


Figure 5.21 The result of nuclei quantification for each nuclear case.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

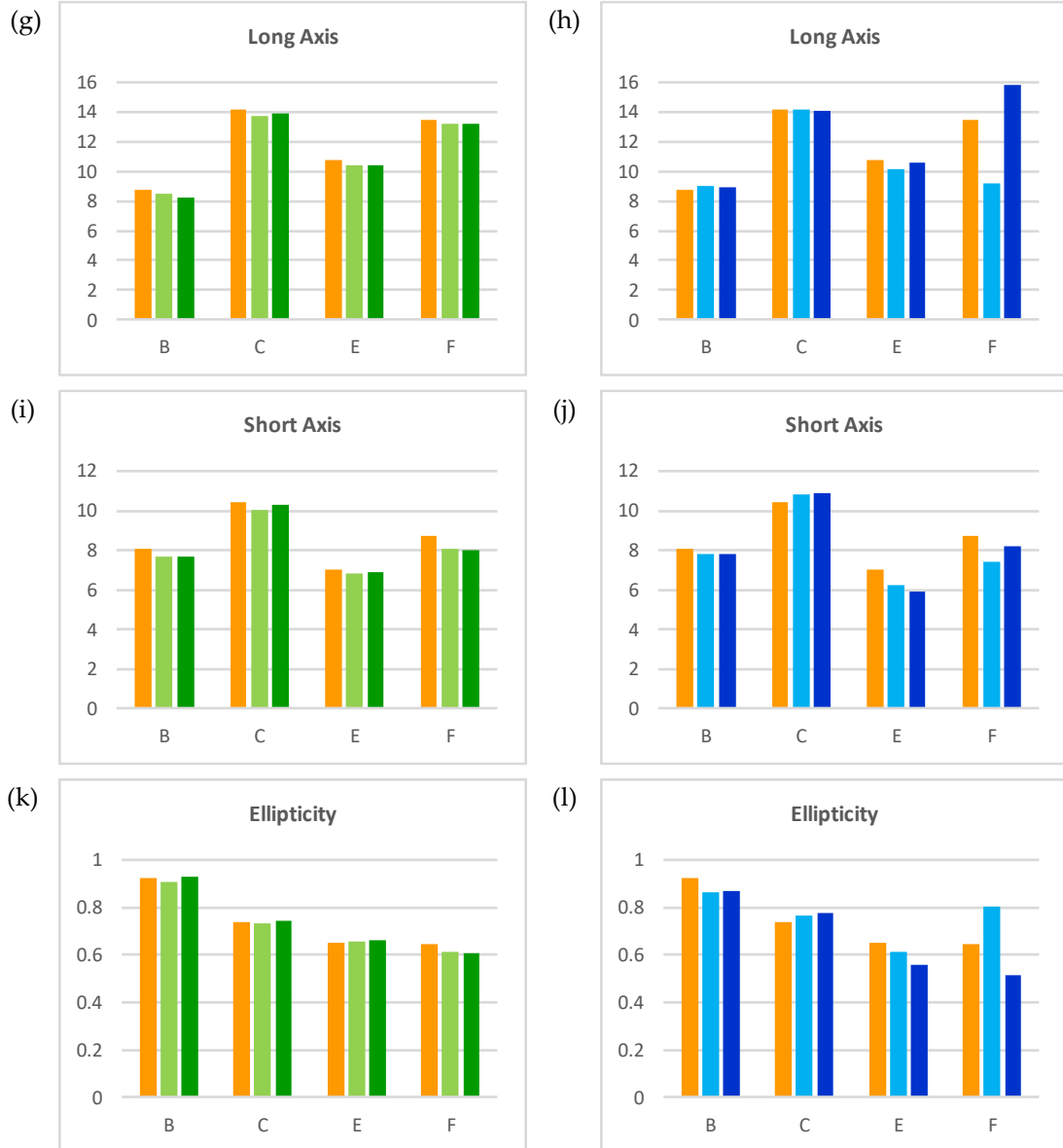


Figure 5.21. (continued).

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

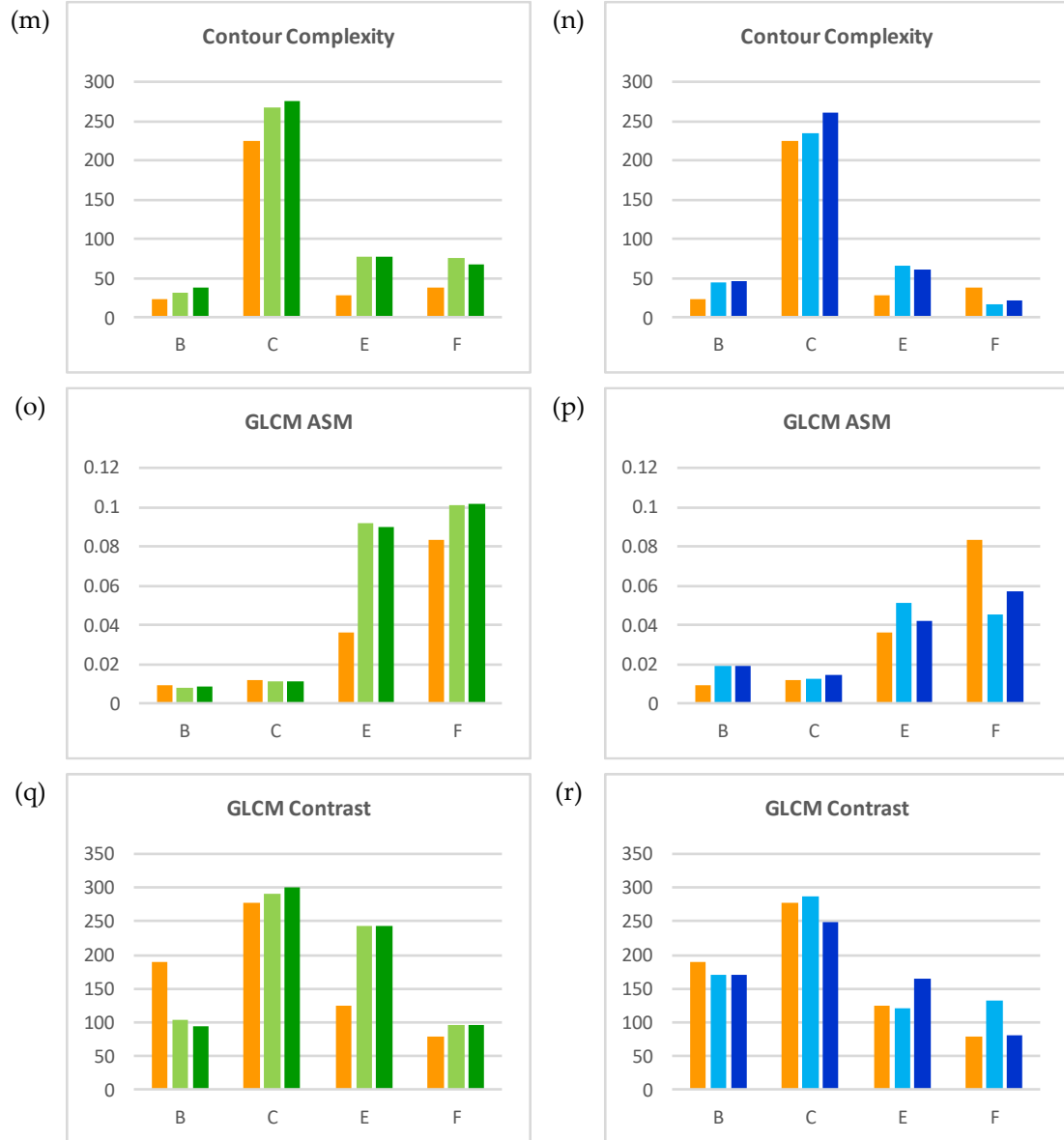


Figure 5.21. (continued).

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

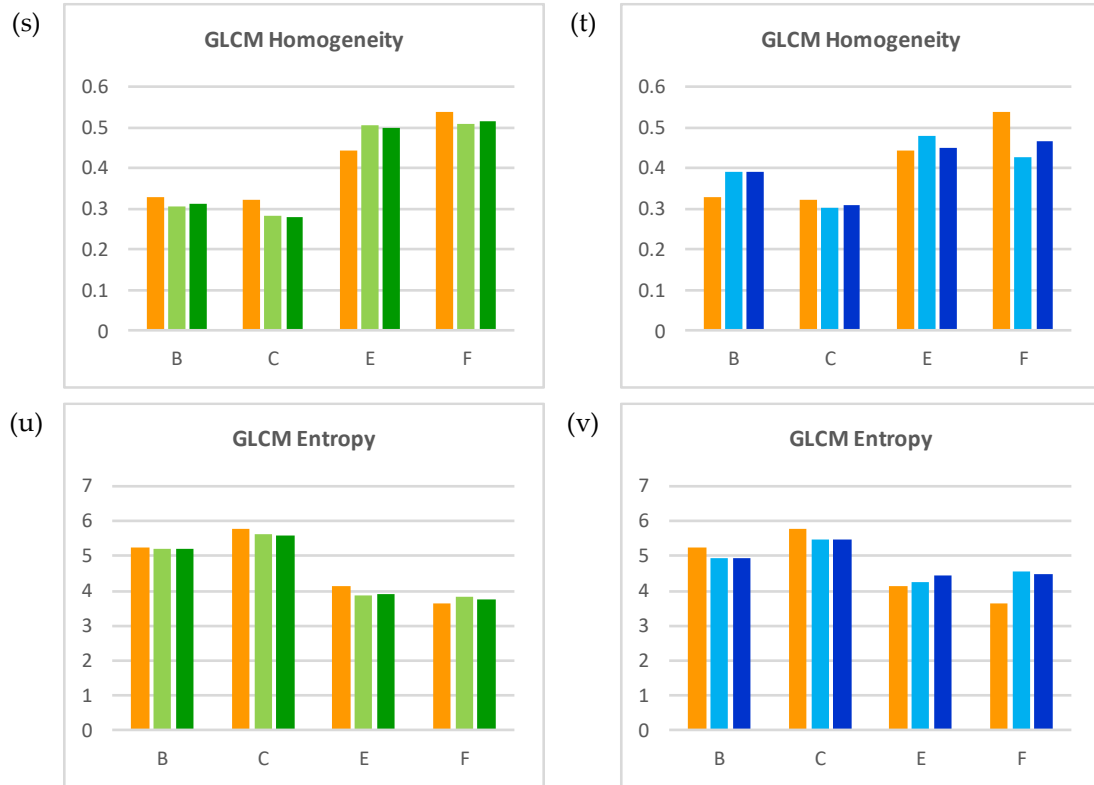


Figure 5.21. (continued).

Table 5.2 Differences on the feature of nuclear B on images without and with sharpness correction where the feature from WSI-A becomes the reference.

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	-19.553	-23.433	↓	-0.566	-2.953	↓
2	Perimeter	-2.349	-2.876	↓	0.081	-0.225	↓
3	Circularity	-0.001	0.000	↑	-0.006	-0.005	↑
4	Long Axis	-0.315	-0.500	↓	0.282	0.202	↓
5	Short Axis	-0.436	-0.415	↑	-0.272	-0.287	↓
6	Ellipticity	-0.017	0.006	↑	-0.059	-0.053	↑
7	Contour Complexity	7.281	14.040	↑	21.656	22.875	↑
8	GLCM ASM	-0.002	-0.001	↑	0.010	0.010	↓
9	GLCM Contrast	-85.716	-95.157	↓	-19.502	-18.681	↑
10	GLCM Homogeneity	-0.022	-0.015	↑	0.062	0.063	↑
11	GLCM Entropy	-0.029	-0.031	↓	-0.306	-0.317	↓

Table 5.3 Differences on the feature of nuclear C on images without and with sharpness correction where the feature from WSI-A becomes the reference.

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	-7.831	-12.275	↓	11.913	11.369	↓
2	Perimeter	-1.521	-1.369	↑	-1.116	-1.256	↓
3	Circularity	0.020	0.007	↓	0.050	0.053	↑
4	Long Axis	-0.453	-0.297	↑	-0.021	-0.118	↓
5	Short Axis	-0.421	-0.124	↑	0.412	0.474	↑
6	Ellipticity	-0.006	0.007	↑	0.030	0.040	↑
7	Contour Complexity	42.122	49.932	↑	9.705	35.965	↑
8	GLCM ASM	-0.001	-0.001	↑	0.001	0.003	↑
9	GLCM Contrast	12.278	22.322	↑	8.522	-29.027	↓
10	GLCM Homogeneity	-0.039	-0.042	↓	-0.021	-0.015	↑
11	GLCM Entropy	-0.153	-0.179	↓	-0.282	-0.294	↓

Table 5.4 Differences on the feature of nuclear E on images without and with sharpness correction where the feature from WSI-A becomes the reference.

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	-12.136	-11.372	↑	-36.187	-9.044	↑
2	Perimeter	-1.562	-1.539	↑	-3.976	0.892	↑
3	Circularity	0.003	0.006	↑	-0.019	-0.064	↓
4	Long Axis	-0.338	-0.364	↓	-0.551	-0.172	↑
5	Short Axis	-0.145	-0.104	↑	-0.753	-1.076	↓
6	Ellipticity	0.007	0.013	↑	-0.039	-0.091	↓
7	Contour Complexity	48.194	48.297	↑	36.538	31.297	↓
8	GLCM ASM	0.056	0.054	↓	0.016	0.006	↓
9	GLCM Contrast	118.295	117.956	↓	-3.828	41.159	↑
10	GLCM Homogeneity	0.063	0.054	↓	0.035	0.007	↓
11	GLCM Entropy	-0.262	-0.242	↑	0.093	0.287	↑

Table 5.5 Differences on the feature of nuclear F on images without and with sharpness correction where the feature from WSI-A becomes the reference.

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	-77.624	-78.783	↓	-190.350	65.585	↑
2	Perimeter	-8.365	-8.414	↓	-19.711	8.872	↑
3	Circularity	0.022	0.020	↓	-0.019	-0.063	↓
4	Long Axis	-0.263	-0.248	↑	-4.247	2.369	↑
5	Short Axis	-0.626	-0.663	↓	-1.306	-0.514	↑
6	Ellipticity	-0.035	-0.038	↓	0.156	-0.129	↓
7	Contour Complexity	37.545	29.188	↓	-21.610	-15.471	↑
8	GLCM ASM	0.018	0.018	↑	-0.038	-0.026	↑
9	GLCM Contrast	18.718	17.682	↓	54.599	2.850	↓
10	GLCM Homogeneity	-0.029	-0.024	↑	-0.111	-0.071	↑
11	GLCM Entropy	0.185	0.098	↓	0.915	0.827	↓

The effect of sharpness correction on statistical distribution of nuclei features on images G and K are presented in Figure 5.22. The 50-percentile values on each nuclear feature extracted from sharpness-corrected images are compared with images corrected with the reference images, i.e. the stain-based color corrected images acquired by WSS-A. The WSI-A2, WSI-B2, and WSI-C2 labels refer to stain-based color-corrected ROI images captured by WSS-A, WSS-B, and WSS-C, respectively, while WSI-B3 and WSI-C3 labels are for stain-based color-corrected images with sharpness correction, captured by WSS-B and WSS-C, respectively. The graphs indicate that there are only limited changes in almost all certain features on both cases, except circularity and ratio of short-long axes having noticeable changes. Table 5.6 and Table 5.7 show the summaries of differences on the quantified nuclear features on image G and K, respectively. Between the four cases, the effect of sharpness correction gave the best effect on image K acquired by WSS-C. Nevertheless, the changes on all feature values are relatively small in all of the cases.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

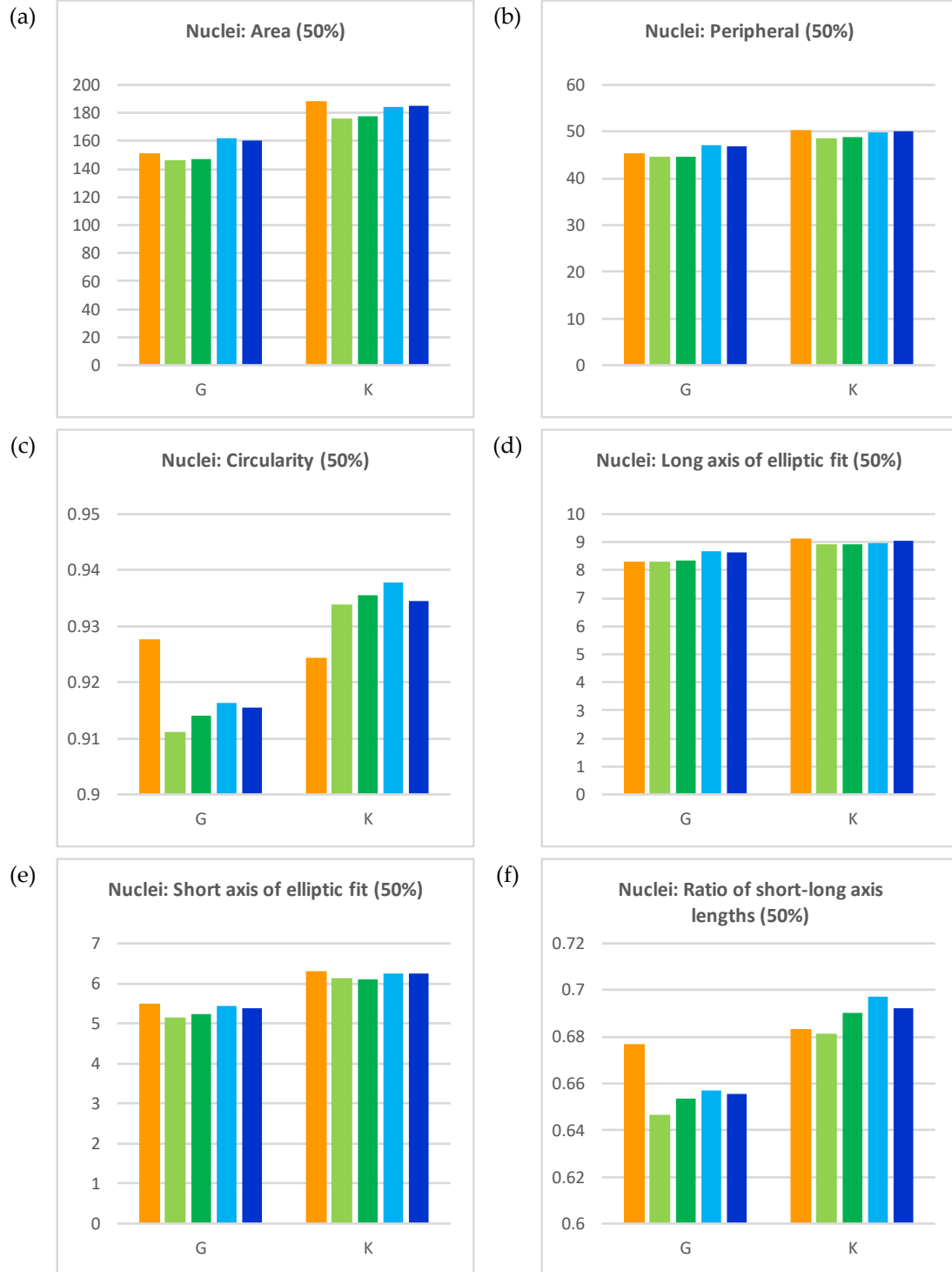


Figure 5.22 The 50-percentile values of three nuclei features on each case.

- WSI-A2 : Stain-based color corrected ROI images acquired by WSS-A
- WSI-B2 : Stain-based color corrected ROI images acquired by WSS-B
- WSI-C2 : Stain-based color corrected ROI images acquired by WSS-C
- WSI-B3 : Stain-based color corrected ROI images acquired by WSS-B with sharpness correction
- WSI-C3 : Stain-based color corrected ROI images acquired by WSS-C with sharpness correction

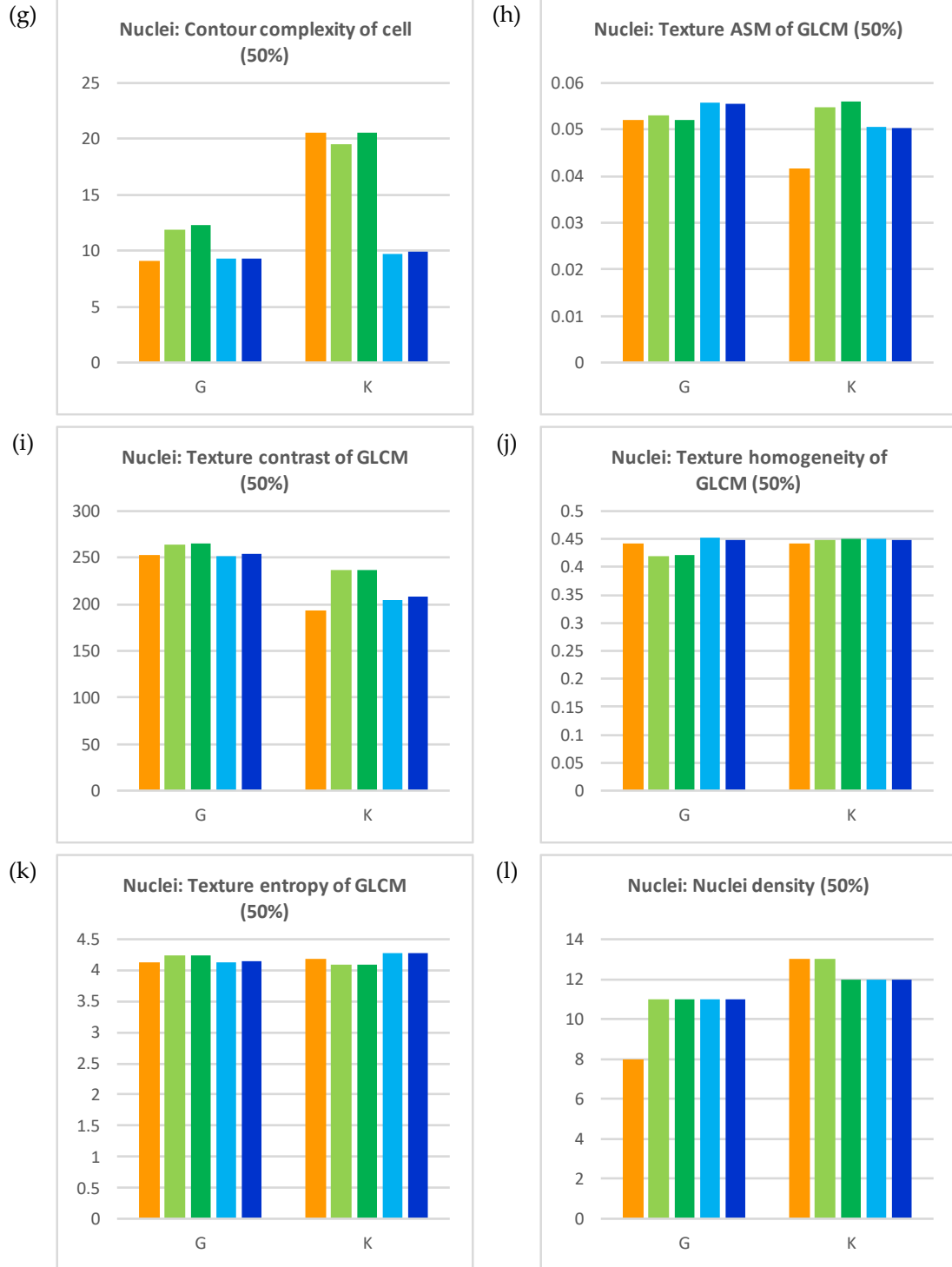


Figure 5.22. (continued).

Table 5.6 Differences on the nuclear feature extracted from image G before and after sharpness correction with the nuclear feature on reference image (WSI-A).

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	5.040	3.978	↓	10.885	9.504	↓
2	Peripheral	0.579	0.612	↑	1.821	1.483	↓
3	Circularity	0.017	0.014	↓	0.011	0.012	↑
4	Long axis of elliptic fit	0.004	0.036	↑	0.363	0.315	↓
5	Short axis of elliptic fit	0.357	0.280	↓	0.073	0.115	↑
6	Ratio of short-long axis lengths	0.030	0.024	↓	0.020	0.021	↑
7	Contour complexity of cell	2.689	3.135	↑	0.141	0.161	↑
8	Texture ASM of GLCM	0.001	0.000	↓	0.004	0.003	↓
9	Texture contrast of GLCM	10.488	11.783	↑	1.229	0.671	↓
10	Texture homogeneity of GLCM	0.023	0.022	↓	0.009	0.005	↓
11	Texture entropy of GLCM	0.103	0.115	↑	0.005	0.021	↑
12	Nuclei density	3	3	n/a	3	3	n/a

Table 5.7 Differences on the nuclear feature extracted from image K before and after sharpness correction with the nuclear feature on reference image (WSI-A).

No	Nuclear Feature (50%)	WSI-B			WSI-C		
		$\Delta B2A2$	$\Delta B3A2$	Diff.	$\Delta C2A2$	$\Delta C3A2$	Diff.
1	Area	12.184	11.004	↓	4.183	3.650	↓
2	Peripheral	1.547	1.305	↓	0.489	0.145	↓
3	Circularity	0.009	0.011	↑	0.013	0.010	↓
4	Long axis of elliptic fit	0.208	0.209	↑	0.134	0.075	↓
5	Short axis of elliptic fit	0.188	0.200	↑	0.074	0.049	↓
6	Ratio of short-long axis lengths	0.002	0.007	↑	0.014	0.009	↓
7	Contour complexity of cell	1.014	0.007	↓	10.802	10.608	↓
8	Texture ASM of GLCM	0.013	0.014	↑	0.009	0.009	↓
9	Texture contrast of GLCM	42.594	43.023	↑	10.573	14.698	↑
10	Texture homogeneity of GLCM	0.005	0.007	↑	0.007	0.005	↓
11	Texture entropy of GLCM	0.100	0.098	↓	0.082	0.092	↑
12	Nuclei density	0	1	↑	1	1	n/a

5.4. Discussion

Image quality plays essential factor in quantitative pathology, especially when image analysis techniques are applied. This chapter is focused on the image quality related to image sharpness. Here, image sharpness is enhanced by deconvoluting the image with Gaussian point spread function, where its size (sigma value) is determined manually through manual observation on each image case. The results indicate that there are less effects of sharpness correction on the intensity profiles and the statistical distribution of nuclei features. Nevertheless, in the basis of single nuclear quantification, some positive results can be founded. This show that the sharpness correction has potential in improving the quality of feature quantification. Further study can be performed on more nuclei. In addition, a method to automatically find the appropriate Gaussian size for each image will be a future work.

Chapter 6

Summary and future work

6.1. Summary

6.1.1. Summary of the proposed methods

There are three main topics in this thesis. In the first part, a method to improve the classification system of histopathological image is proposed. Based on the evaluation on the effect of image masking, comprehending structural features with nuclei features, and using a set of biopsy and surgical resected samples as training data, the classification rates can be improved and a more reliable system can be obtained. The masking process improve the reliability of nuclei features as falsely detected nuclei are removed from the quantification. The masking process also enable the extraction of new set of structural features. Combination of the nuclei features with trabecular and tissue changes features improves the classification rate in in the case of HCC detection using SVM, especially in the cases of lower grade HCC. Based on these methods, an approach to enhance automatic classification of cancer on histopathological images was proposed. On the other hand, the study on histopathological image classification based on the tissue type ahead of the cancer classification has potential in improving the cancer detections. Yet, further study can be carried out with larger data size on this aspect.

In the second part, a series of evaluations were carried out to evaluate the effectiveness of three color correction methods on quantitative pathology. Three color correction methods are stain-based color correction, device-based, and a combination of device- and stain-based color correction. Each color correction method was developed on to correct the different causes of color variation on WSI. Based on the evaluations, the stain-based color correction gives better impact compared to the device-based color correction. The stain-based method can even compensate the color difference due to the device varieties. Color correction is essential in the process of feature quantifications of histopathological images, especially in order to gain reliable nuclei morphological features. The study also shows that the texture related features of nuclear are more sensitive to image contrast, compared to the morphological features. Nevertheless, some

difference on the quantified features are still existed. The difference may be caused from the difference on image quality due to the image sharpness.

The last topic focused on the sharpness correction on histopathological images and its effects on quantitative pathology. In the first part, an MTF-based image correction is introduced. Nonetheless, it lacks of effectiveness since the method derives MTF from histopathological images, while the image may have different focus every time it is scanned. In the second part, sharpness correction by utilizing Gaussian point spread function is proposed. Based on the evaluation on the method on the intensity profiles and the quantified nuclear features, the method has potential to improve the reliability of the features. The study also shows that the morphological related nuclear features are more robust and more independent to device compared to the texture related nuclear features.

6.1.2. Overall concept of device-independent histological quantification and automatic classification system

Figure 6.1 illustrates the overall concept of device-independent histological quantification and automatic classification system. Generally, the system will consist on three main components; (1) the image preprocessing component, (2) tissue structure analysis component, and (3) classification component.

Three general types of WSSs are included in the figure to represent the reference scanner (WSS-A), different scanners that have similar output format with the reference scanner (WSS-B), and scanners that produce different output format with the reference scanner (WSS-C). The image processing component will receive the input WSIs, perform image conversion and sampling adjustment when necessary, and generate ROI images for further processing. Digital corrections which consist of sharpness correction and stain-based color correction are then performed on the ROI images. Here, device-based color correction is not implemented since the stain-based color correction can compensate the color difference due to the device varieties

The tissue structure analysis focuses on segmenting the tissue components and extracting features from the tissue. In the proposed system, four sets of features will be quantified, consisting nuclear features, trabecular features, tissue changes features, and tissue texture features.

The classification component will perform classifications on the histopathological images. First, classification on the tissue types will be performed based on the tissue texture features. The tissue type will determine the set of features used, whether based on nuclear features only or based on combination of nuclear and tissue structural features, for cancer classification.

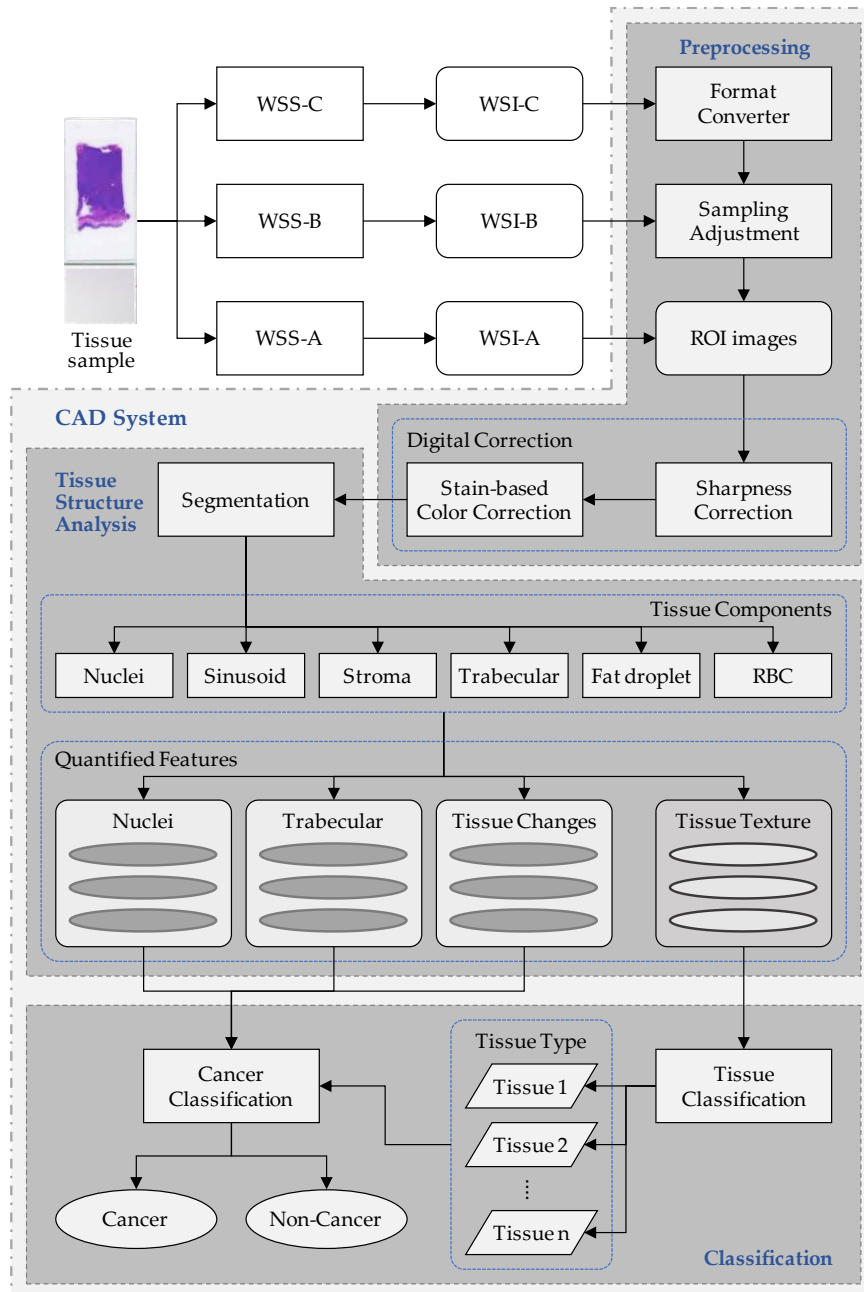


Figure 6.1 The concept of device-independent histological quantification and automatic classification system.

6.2. Future Work

A lot of works can be performed in the future in order to gain a device-independent histological quantification and automatic classification system. Mainly, the works can be categorized into two main fields; enhancing the classification rates and exploring more options to reduce the feature quantification on device dependence. On the first field, one of the focus will be developing method to classify histopathological images based on the tissue type with larger data set. By classifying the image based on the tissue type, appropriate set of features suitable for cancer detection can be determined. Further exploration on the tissue components and characteristics may lead into more features to be quantified. Other work that can be addressed is to improve the system's robustness. By developing algorithms that are more robust in quantifying the features compared to the current algorithm.

On the other field, developing a suitable color chart to calibrate and correct color difference due to device used to acquire the histopathological images is important. Studies on digital staining on tissue may also help improving the quality of color on histopathological images. The effect of color correction on pathological diagnosis by pathologists on digitized slide may also be carried out so that a better analysis on the digitized samples. Meanwhile, to improve image quality regarding sharpness difference, a method to automatically calculate the appropriate Gaussian size of the point spread function used for image deconvolution can be carried out. Developing appropriate slide to derive the MTF of WSS is also necessary to characterize the imaging system better.

Publications and Conferences

1. Journal papers (referred)

- [1] **MA Aziz**, H Kanazawa, Y Murakami, et al. Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. *J Pathol Inform.* 2015; 6: 26. <https://doi.org/10.4103/2153-3539.158044>
- [2] **MA Aziz**, T Nakamura, M Yamaguchi, et al. Effectiveness of color correction on quantitative analysis of histopathological images acquired by different whole slide scanners. *Artif Life Robotics* (2018). <https://doi.org/10.1007/s10015-018-0451-0>

2. International conference

- [1] **MA Aziz**, H Kanazawa, Y Murakami, et al. Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. 2nd Congress of the International Academy of Digital Pathology, Harvard Medical School, Boston, Massachusetts, November 3-7, 2014

3. Journal papers (Non-referred)

- [1] **MA Aziz**, H Kanazawa, Y Murakami, et al. Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features. *Anal Cell Pathol.* 2014; 2014: 726782. [Meeting Abstract]

4. Domestic conferences

- [1] M Ishikawa, Y Fujita, **MA Aziz**, et al. Examination of improve the accuracy of HCC classification by quantification of tissue structure from HE-Stained liver specimens. *Proc. of JAMIT Annual Meeting 2014*; 33:1-9 (ID OP3-7)
- [2] **MA Aziz**, T Nakamura, M Yamaguchi. Effectiveness of color correction on quantitative analysis of histopathological images acquired by different whole slide scanners. The 5th Symposium of the 'Color' of Digital Imaging in Biomedicine, Chiba University, Chiba, February 15, 2018

Bibliography

- [1] World Health Organization, "WHO | NCD mortality and morbidity," WHO. World Health Organization, 2017.
- [2] International Agency for Research on Cancer, *World Cancer Report 2014*. Lyon, 2014.
- [3] International Agency for Research on Cancer, "Simple Maps." [Online]. Available: <http://globocan.iarc.fr/Pages/Map.aspx>. [Accessed: 22-Jun-2018].
- [4] C.-W. Wang and C.-P. Yu, "Automatic Morphological Classification of Lung Cancer Subtypes with Boosting Algorithms for Optimizing Therapy," in *Machine Learning in Medical Imaging*, K. Suzuki, F. Wang, D. Shen, and P. Yan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 217–224.
- [5] C. Nunes *et al.*, "High agreement between whole slide imaging and optical microscopy for assessment of HER2 expression in breast cancer: Whole slide imaging for the assessment of HER2 expression," *Pathol. - Res. Pract.*, vol. 210, no. 11, pp. 713–718, Nov. 2014.
- [6] V. Ojansivu *et al.*, "Automated classification of breast cancer morphology in histopathological images," *Diagnostic Pathology*, vol. 8, no. Suppl 1. p. S29, 2013.
- [7] S. Doyle, M. D. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi, "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinformatics*, vol. 13, no. 1, p. 282, Oct. 2012.
- [8] A. Saito, E. Cosatto, T. Kiyuna, and M. Sakamoto, "Dawn of the digital diagnosis assisting system, can it open a new age for pathology?," *Proc.SPIE*, vol. 8676, pp. 8676-8676–16, 2013.
- [9] Y. Sucaet, W. Waelput, P. Vermeulen, and G. Van Den Eynden, "Using a Novel Whole Slide Imaging Software Platform for an International Multicenter Validation Study to Assess the Histological Growth Pattern of Liver Metastases," *Anal. Cell. Pathol.*, vol. 2014, pp. 2–4, Dec. 2014.
- [10] R. Randell *et al.*, "Effect of Display Resolution on Time to Diagnosis with Virtual Pathology Slides in a Systematic Search Task," *J. Digit. Imaging*, vol. 28, no. 1, pp. 68–76, 2014.
- [11] E. Cosatto *et al.*, "Automated gastric cancer diagnosis on H&E-stained sections; ltraining a classifier on a large scale with multiple instance machine learning," in *Proc.SPIE*, 2013, vol. 8676, pp. 8676-8676–9.
- [12] M. Ishikawa *et al.*, "HE染色肝病理標本における組織構造定量化によるがん検出精度向上の検討 [Examination of improve the accuracy of HCC classification by quantification of tissue structure from HE-stained liver specimens]," in *日本医用画像工学会大会予稿集(CD-ROM) [Proc. of JAMIT Annual Meeting]*, 2014, vol. 33, pp. OP3-7.
- [13] A. S. Jørgensen *et al.*, "Using cell nuclei features to detect colon cancer tissue in hematoxylin and eosin stained slides," *Cytom. Part A*, vol. 91, no. 8, pp. 785–793,

- 2017.
- [14] T. Kiyuna *et al.*, "Automatic classification of hepatocellular carcinoma images based on nuclear and structural features," *Proc.SPIE*, vol. 8676, pp. 8676-8676-6, 2013.
 - [15] M. K. K. Niazi *et al.*, "Visually Meaningful Histopathological Features for Automatic Grading of Prostate Cancer," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 4, pp. 1027-1038, Jul. 2017.
 - [16] D. Romo-Bucheli, A. Janowczyk, H. Gilmore, E. Romero, and A. Madabhushi, "Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images," *Sci. Rep.*, vol. 6, p. 32706, Sep. 2016.
 - [17] M. A. Aziz *et al.*, "Enhancing automatic classification of hepatocellular carcinoma images through image masking, tissue changes and trabecular features," *J. Pathol. Inform.*, vol. 6, no. 1, p. 26, 2015.
 - [18] Y. Yamashita *et al.*, "Development of a Prototype for Hepatocellular Carcinoma Classification Based on Morphological Features Automatically Measured in Whole Slide Images," *Analytical Cellular Pathology*, vol. 2014, p. 2, 2014.
 - [19] M. Valkonen, K. Kartasalo, K. Liimatainen, M. Nykter, L. Latonen, and P. Ruusuvaori, "Metastasis detection from whole slide images using local features and random forests," *Cytom. Part A*, vol. 91, no. 6, pp. 555-565, 2017.
 - [20] B. Têtu and A. Evans, "Canadian Licensure for the Use of Digital Pathology for Routine Diagnoses: One More Step Toward a New Era of Pathology Practice Without Borders," *Arch. Pathol. Lab. Med.*, vol. 138, no. 3, pp. 302-304, Mar. 2014.
 - [21] J. Griffin and D. Treanor, "Digital pathology in clinical use: Where are we now and what is holding us back?," *Histopathology*, vol. 70, no. 1, pp. 134-145, 2017.
 - [22] American Society of Clinical Oncology, "Reading a Pathology Report | Cancer.Net," 2016. [Online]. Available: <https://www.cancer.net/navigating-cancer-care/diagnosing-cancer/reports-and-results/reading-pathology-report>. [Accessed: 27-Jun-2018].
 - [23] American Society of Clinical Oncology, "After a Biopsy: Making the Diagnosis | Cancer.Net," 2018. [Online]. Available: <https://www.cancer.net/navigating-cancer-care/diagnosing-cancer/reports-and-results/after-biopsy-making-diagnosis>. [Accessed: 27-Jun-2018].
 - [24] D. Treanor, "Virtual slides: an introduction," *Diagnostic Histopathol.*, vol. 15, no. 2, pp. 99-103, 2009.
 - [25] A. Buhmeida, "Quantitative Pathology: Historical Background, Clinical Research and Application of Nuclear Morphometry and DNA Image Cytometry," *Libyan J. Med.*, vol. 1, no. 2, pp. 126-139, 2006.
 - [26] P. H. Bartels, "Future directions in quantitative pathology: Digital knowledge in diagnostic pathology," *J. Clin. Pathol.*, vol. 53, no. 1, pp. 31-37, 2000.
 - [27] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review #x2014;Current Status and Future Potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97-114, 2014.

- [28] A. Nagase, M. Takahashi, and M. Nakano, "Automatic calculation and visualization of nuclear density in whole slide images of hepatic histological sections," *Biomed. Mater. Eng.*, vol. 26, no. s1, pp. S1335–S1344, Aug. 2015.
- [29] S. Wienert *et al.*, "Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach," *Sci. Rep.*, vol. 2, pp. 1–7, 2012.
- [30] M. Veta *et al.*, "Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer," *Mod. Pathol.*, vol. 25, no. 12, pp. 1559–1565, 2012.
- [31] S. Wen, T. M. Kurc, Y. Goa, T. Zhao, J. Saltz, and W. Zhu, "A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image," *J. Pathol. Inform.*, vol. 8, no. 1, p. 38, 2017.
- [32] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [33] M. Ishikawa *et al.*, "Automatic segmentation of hepatocellular structure from HE-stained liver tissue," *Proc. SPIE*, vol. 8676, pp. 8676–8676–7, 2013.
- [34] M. Ishikawa *et al.*, "Segmentation of Sinusoids in Hematoxylin and Eosin Stained Liver Specimens Using an Orientation-Selective Filter," *Open J. Med. Imaging*, vol. 03, no. 04, pp. 144–155, 2013.
- [35] M. Ishikawa *et al.*, "HE染色肝病理組織標本における索状構造の配列の乱れ定量化法の提案 [A method to disarray quantification of trabeculae from HE-stained liver specimens]," in *日本医用画像工学会大会予稿集(CD-ROM) [Proc. of JAMIT Annual Meeting]*, 2013, vol. 32, pp. OP7-5.
- [36] M. Ishikawa *et al.*, "肝病理組織における索状構造解析のための形態的特徴の自動計測 [An automatic measurement of morphological feature for analysis of the hepatic trabeculae]," *信学技報 [IEICE Technical Report]*, vol. 112, no. 411, pp. 255–260, 2013.
- [37] M. Ishikawa *et al.*, "機械学習を用いた肝生検病理画像からの脂肪滴抽出のための一手法 [A method to extract fat drops from liver biopsy image using machine learning algorithm]," *信学技報 [IEICE Technical Report]*, vol. 113, no. 410, pp. 103–108, 2014.
- [38] M. Ishikawa *et al.*, "Automatic quantification of morphological features for hepatic trabeculae analysis in stained liver specimens," *J. Med. Imaging*, vol. 3, pp. 3–3–13, 2016.
- [39] H. Komagata *et al.*, "肝病理組織画像における核のグラフ構造特徴量高速計算アルゴリズム [High speed algorithm to calculate feature quantities of nuclei graph structures in hepatic histological images]," *信学技報 [IEICE Tech. Report]*, vol. 113, no. 410, pp. 239–244, 2014.
- [40] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological Image Analysis: A Review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [41] A. Cruz-Roa *et al.*, "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, p. 46450, Apr. 2017.

- [42] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Inform.*, vol. 7, no. 1, p. 29, 2016.
- [43] International Agency for Research on Cancer, "Liver Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012." [Online]. Available: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx. [Accessed: 04-Jul-2015].
- [44] F. T. Bosman, F. Carneiro, R. H. Hruban, and N. D. Theise, *WHO Classification of Tumours of the Digestive System*, 4th ed. Lyon: IARC Press, 2010.
- [45] R. B. Birrer, D. Birrer, and J. V. Klavins, "Review: Hepatocellular carcinoma and hepatitis virus," *Ann. Clin. Lab. Sci.*, vol. 33, no. 1, pp. 39–54, 2003.
- [46] A. M. Di Bisceglie, "Hepatitis B and hepatocellular carcinoma," *Hepatology*, vol. 49, no. S5, pp. S56–S60, 2009.
- [47] E. S. Bialecki and A. M. Di Bisceglie, "Diagnosis of hepatocellular carcinoma," *HPB*, vol. 7, no. 1, pp. 26–34, Mar. 2005.
- [48] M. Pazgan-Simon, S. Serafinska, J. Janocha-Litwin, K. Simon, and J. Zuwała-Jagiello, "Diagnostic challenges in primary hepatocellular carcinoma: case reports and review of the literature.," *Case Rep. Oncol. Med.*, vol. 2015, p. 5, Apr. 2015.
- [49] S. N. Martins-Filho, C. Paiva, R. S. Azevedo, and V. A. F. Alves, "Histological Grading of Hepatocellular Carcinoma—A Systematic Review of Literature," *Front. Med.*, vol. 4, p. 193, 2017.
- [50] V. Paradis, "Histopathology of Hepatocellular Carcinoma," in *Multidisciplinary Treatment of Hepatocellular Carcinoma*, J.-N. Vauthey and A. Brouquet, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 21–32.
- [51] Y. Murakami *et al.*, "Color processing in pathology image analysis system for liver biopsy," in *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, 2014, vol. 2014, no. January, pp. 184–188.
- [52] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27:1–27:27, May 2011.
- [53] C.-C. Chang and C.-J. Lin, "LIBSVM -- A Library for Support Vector Machines." [Online]. Available: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>. [Accessed: 16-Jul-2018].
- [54] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall, 2008.
- [55] P. A. Bautista, N. Hashimoto, and Y. Yagi, "Color standardization in whole slide imaging using a color calibration slide," *J. Pathol. Inform.*, vol. 5, no. 1, p. 4, 2014.
- [56] E. L. Clarke *et al.*, "Development of a novel tissue-mimicking color calibration slide for digital microscopy," *Color Res. Appl.*, vol. 43, no. 2, pp. 184–197, 2017.
- [57] Y. Murakami *et al.*, "Color correction in whole slide digital pathology," in *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, 2012, pp. 253–258.
- [58] M. A. Aziz *et al.*, "Effectiveness of color correction on the quantitative analysis of histopathological images acquired by different whole-slide scanners," *Artif. Life Robot.*, vol. 0, no. 0, p. 0, 2018.

Appendices

Appendix A

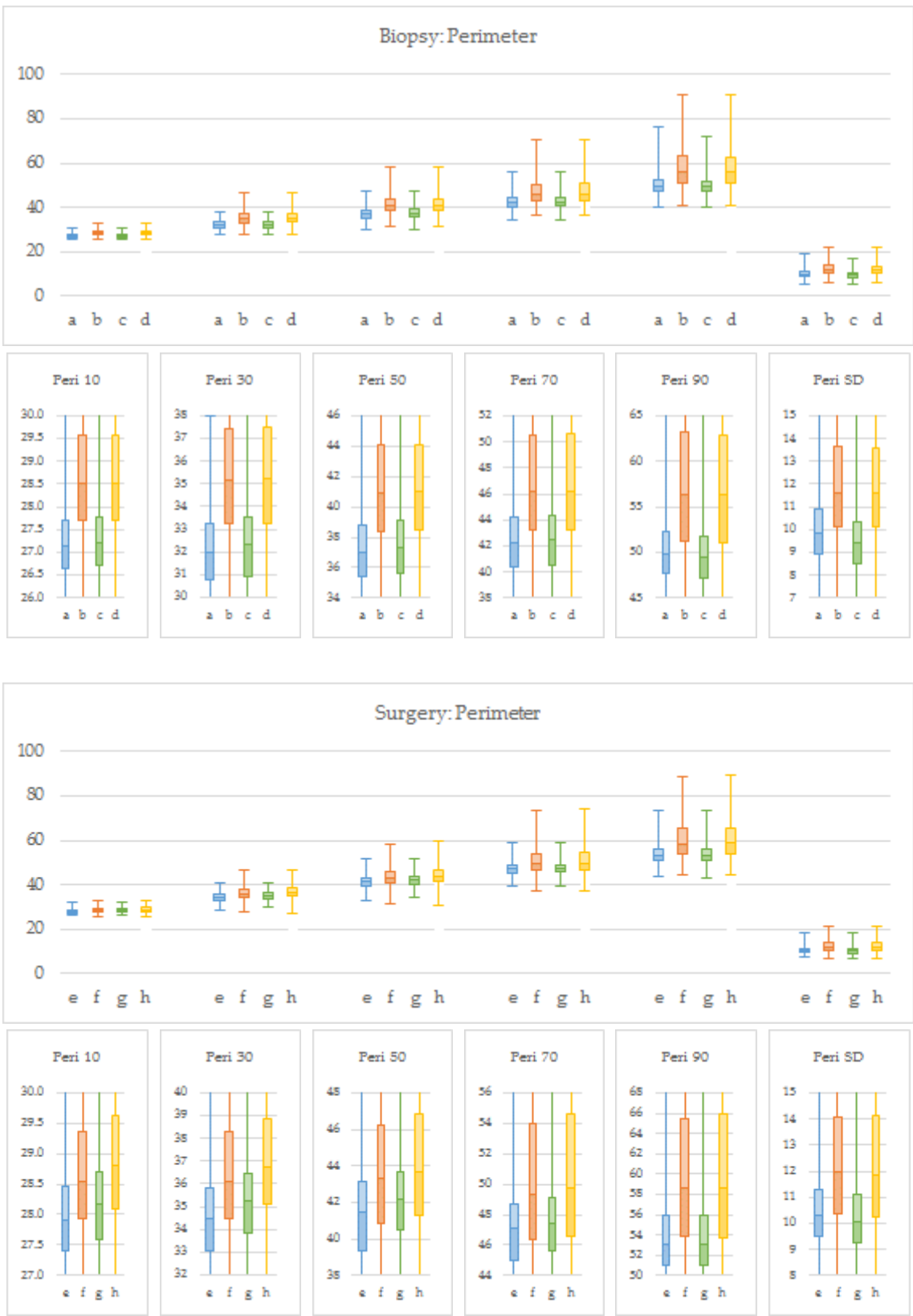
Box-and-whisker plots showing the overall distributions of cancer and non-cancer data on each nuclear feature before and after image masking. Refer to the following table for the details on each plot.

Plot	Sample Type	Cancer (HCC)	Masking process
(a)	Biopsy	No	No
(b)	Biopsy	Yes	No
(c)	Biopsy	No	Yes
(d)	Biopsy	Yes	Yes
(e)	Surgical resected	No	No
(f)	Surgical resected	Yes	No
(g)	Surgical resected	No	Yes
(h)	Surgical resected	Yes	Yes

1. Feature: area



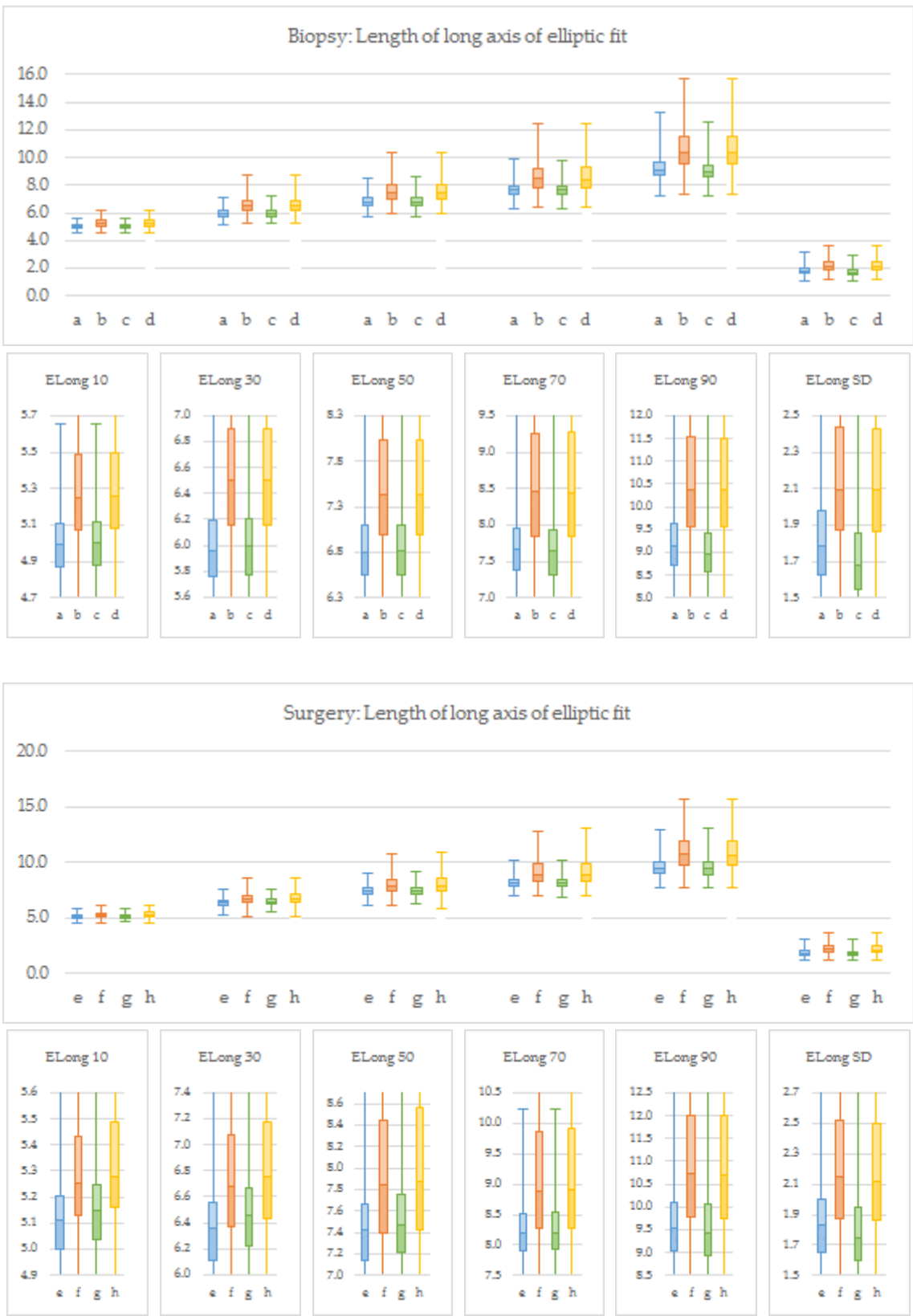
2. Feature: periphery



3. Feature: circularity



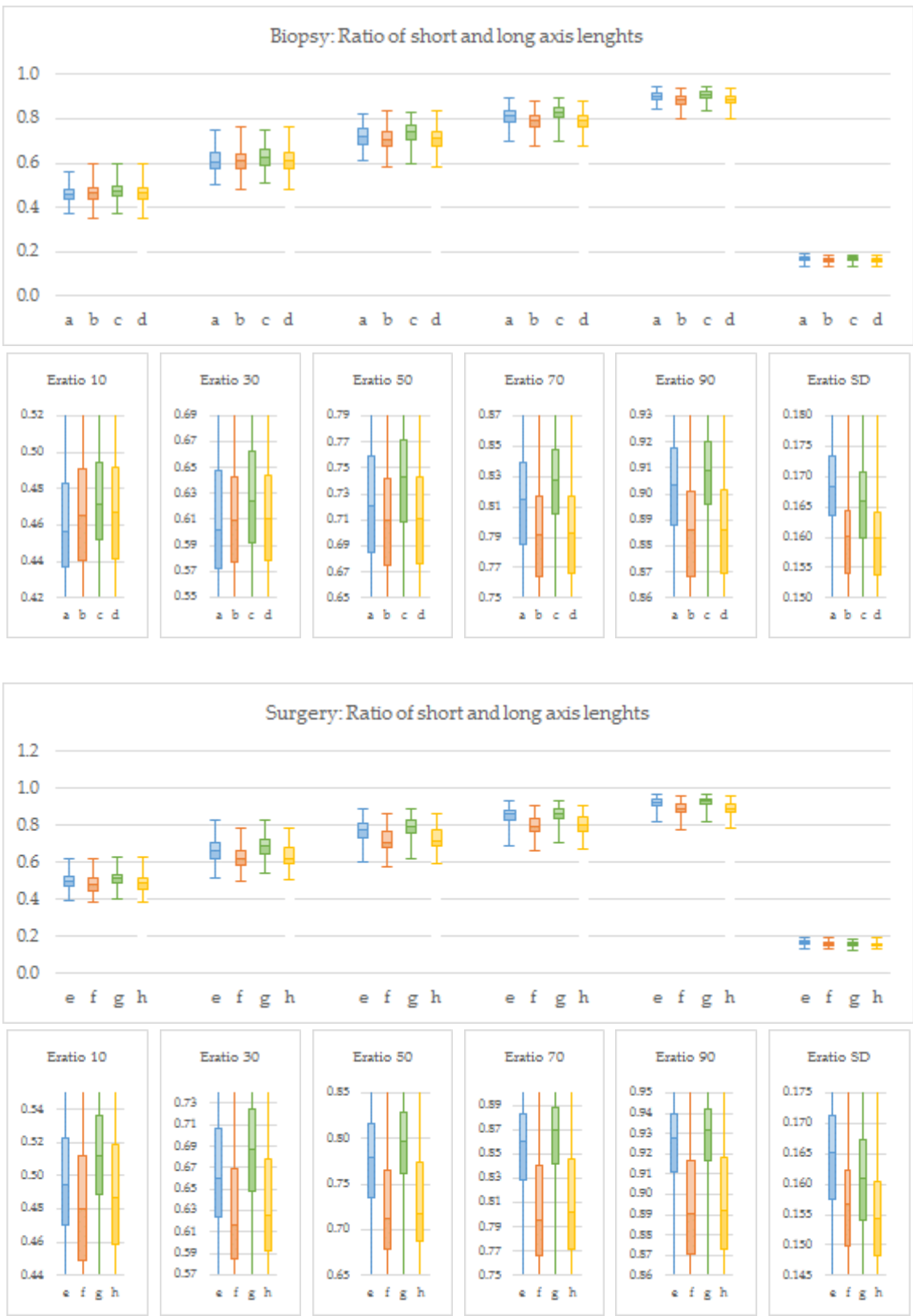
4. Feature: length of long axis of elliptic fit



5. Feature: length of short axis of elliptic fit



6. Feature: ratio of short and long axis lengths



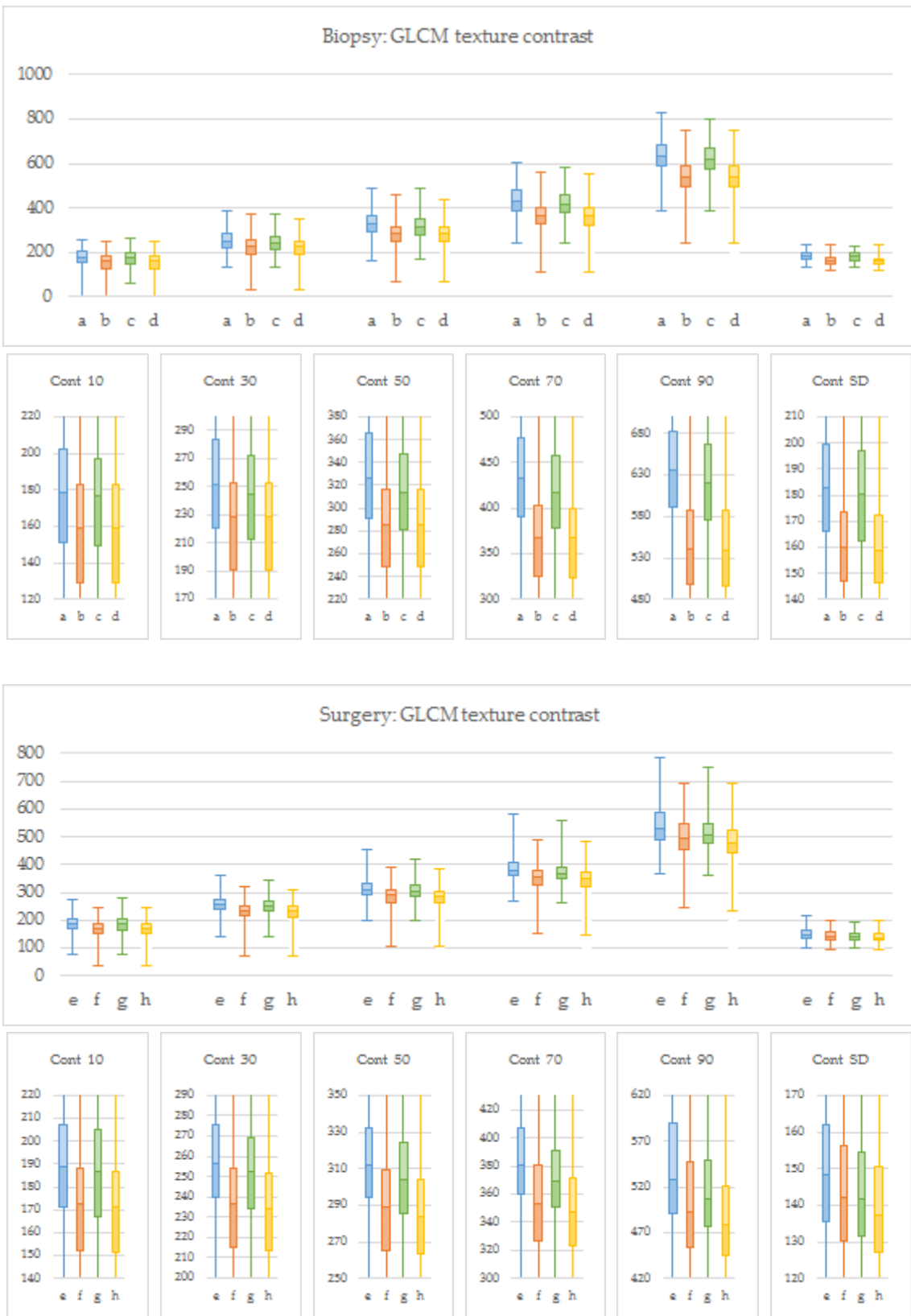
7. Feature: contour complexity of cell



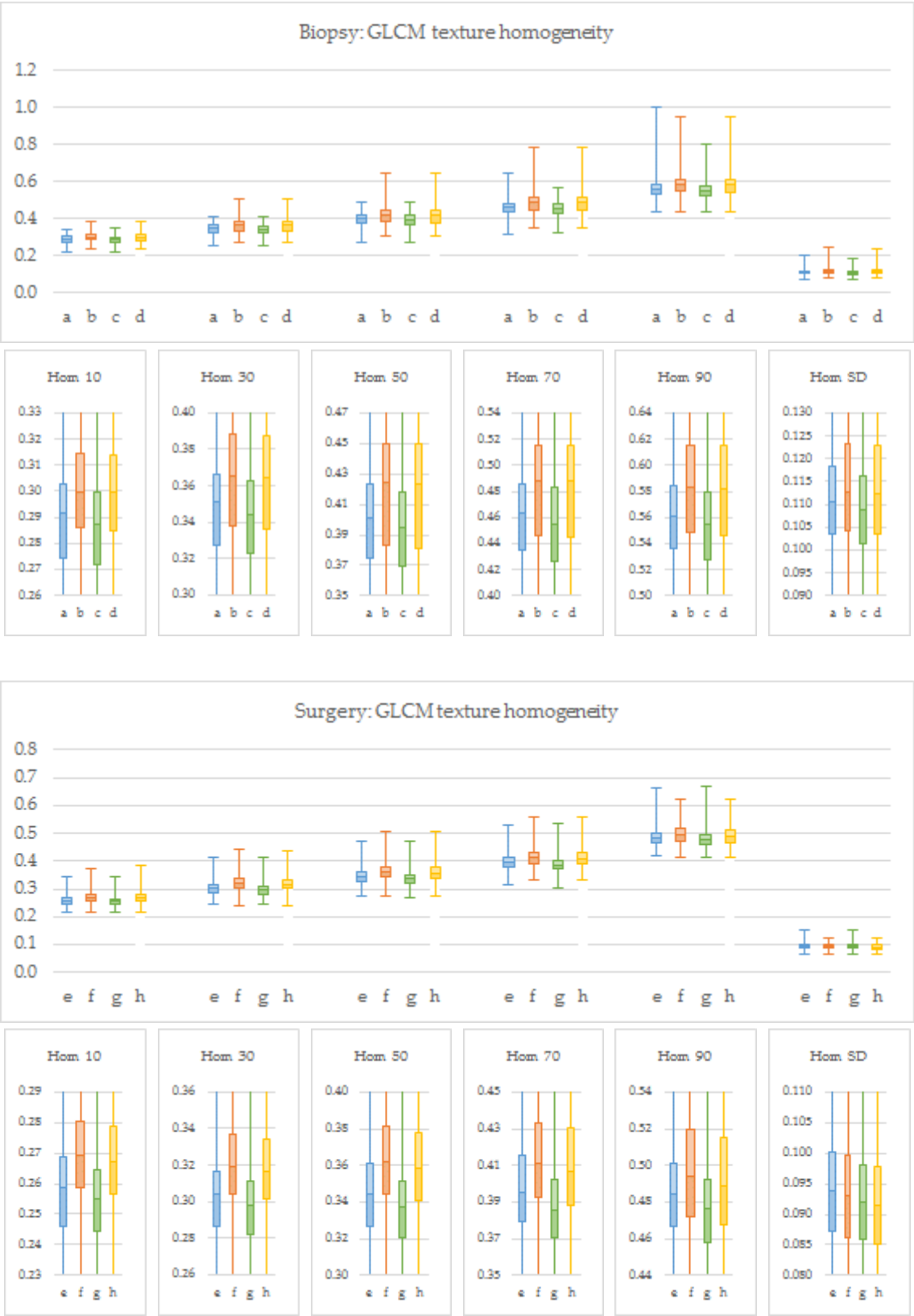
8. Feature: GLCM ASM



9. Feature: GLCM texture contrast



10. Feature: GLCM texture homogeneity



11. Feature: GLCM texture entropy



12. Feature: nuclei density

