



AGH University of Science and Technology
FACULTY OF ELECTRICAL ENGINEERING, AUTOMATICS, COMPUTER SCIENCE
AND BIOMEDICAL ENGINEERING

DEPARTMENT OF AUTOMATIC CONTROL AND ROBOTICS

Ph.D. Thesis

*Computer vision and machine learning algorithms
for dermatopathological image analysis*

Author: *Paweł Kłeczek, M.Sc.*
Discipline: *Automatics, Electronics and Electrical Engineering*
Supervisor: *Joanna Jaworek-Korjakowska, Ph.D., Associate Professor*

Kraków, 2020



Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI
I INŻYNIERII BIOMEDYCZNEJ
KATEDRA AUTOMATYKI I ROBOTYKI

Rozprawa doktorska

*Algorytmy wizji komputerowej i uczenia maszynowego
w analizie obrazów dermatopatologicznych*

Autor: *mgr inż. Paweł Kłeczek*
Dyscyplina: *Automatyka, Elektronika i Elektrotechnika*
Promotor: *dr hab. inż. Joanna Jaworek-Korjakowska, prof. AGH*

Kraków, 2020

Above all I would like to thank my supervisor Joanna Jaworek-Korjakowska, Ph.D., Associate Prof. for her patient guidance, advice she has provided throughout my time as a doctoral student, and constant encouragement to take up new challenges. Her careful editing contributed enormously to the production of this thesis. She is a person I well and truly consider my mentor.

I would like to thank my former supervisor Prof. Ryszard Tadeusiewicz for his care and support during the initial period of my PhD studies.

I also thank Grzegorz Dyduch, M.D., Ph.D. from the Chair of Pathomorphology, Jagiellonian University Medical College for his selfless assistance and guidance on all medical aspects of my researches, particularly on collecting the medical data of skin lesions. Without his help this thesis would never come into existence.

Finally, I am grateful to my family – my parents and my brother – for their continuous support and for mobilizing me during all the years of conducting PhD-related researches and during the preparation of this thesis.

Abstract

Research aim

The main research aim of this thesis was to propose an algorithmic solution, which would make use of advanced image processing and machine learning methods, to the problem of determining the tissue structure of human epidermis in histopathological images of H&E-stained skin specimens, and to develop a complete system for automated diagnostics of selected skin melanocytic lesions based on the analysis of lesion's epidermal morphometry. To the best of Author's knowledge it is the first attempt to automatically determine epidermal morphometry of a lesion and to utilize that information for the diagnostic purpose.

The major and minor theses were formulated as follows:

Main thesis By means of advanced image processing methods and machine learning algorithms it is possible to analyze the tissue structure of human epidermis in histopathological images.

Thesis I Using the proposed methods for image segmentation and image analysis it is possible to assess the degree of malignancy of a given skin lesion.

Thesis II By means of machine learning methods it is possible to formulate the basic histopathological criteria for diagnosing skin lesions.

The proposed solution

The proposed methods and the whole diagnostic system were designed to automatically analyze and distinguish between the three common types of skin melanocytic lesions: lentiginos (benign lesions), dysplastic junctional nevi (continua between benign and malignant lesions) and melanomas (malignant lesions).

The need for robust and efficient (automatic) diagnostic tools in dermatopathology is evident due to a significant and constant increase in the incidence rate of and mortality from cutaneous melanoma, the most aggressive and dangerous skin cancer, among Caucasian populations worldwide. The traditional, manual way of diagnosing skin lesions, the histopathological examination, has three main drawbacks: it is prone to human errors, its results are often non-reproducible, and the analysis of a large volume of specimen is laborious. By applying the methods of computer vision and machine learning to slides digitized at high magnification it is possible to address these issues – to improve diagnostic accuracy, to increase the reproducibility of diagnostic outcomes, as well as to shorten the time needed for making a diagnosis.

In the thesis the following image analysis algorithms were proposed, implemented and verified:

- A tissue segmentation method which utilizes both global and local information about color distribution in the CIELAB color space extracted from a 3D histogram of pixel intensities using statistical and morphological analysis.
- An epidermis segmentation method which utilizes information about shape and distribution of slide background regions in an image as well as information about distribution and concentration of hematoxylin and eosin stains extracted from their joint histogram (using statistical and image processing methods).
- A rete segmentation algorithm which automatically determines the basic morphometric features of individual rete ridges (i.e., their location, base width, length, and height) in a segmented epidermis. It is based mainly on analyzing the curvature of the epidermal boundary to identify endpoints of projection bases, and on analyzing the skeleton of the epidermal region in the graph representation, using graph algorithms and morphological operations, to identify rete bases and the location of retes along the epidermis main axis.

Each of the aforementioned algorithms is fully automated and to a large extent based on domain-specific knowledge of morphometric and biochemical properties of skin tissue structures.

Next, indexes describing the morphometry of the epidermal component, based on information about morphometry and location of individual retes, were presented. They were then used as predictors in the proposed automatic skin lesions classification algorithm – in this thesis four machine learning models were evaluated out of which the support vector machine (SVM) classifier gave the most accurate results. Finally, a statistical analysis of various relationships between epidermal morphometry and basic types of skin melanocytic lesions was carried out by means of a decision tree model in order to propose diagnostic criteria which could be used during histopathological examination.

The performance of the whole system was verified on a dataset consisting of 75 high-resolution WSI images of skin melanocytic lesions (25 for each lesion type) diagnosed by an expert dermatopathologist. It scored the accuracy of 74.7% and 86.7% for the “lentigo vs. junctional dysplastic nevus vs. melanoma” and “benign vs. malignant” classification tasks, respectively. For the same classification tasks a classification tree model (constructed based on the proposed morphometric indexes) scored the accuracy of 73.3% and 88.0%, respectively, and thus could be used to help formulate the basic histopathological criteria related to epidermal morphometry.

In Author’s opinion the obtained results let him pronounce that the main research aim of the proposed project was reached and that both theses were proven.

Streszczenie pracy

Cel pracy

Głównym celem badawczym niniejszej rozprawy było zaproponowanie algorytmicznego rozwiązania wykorzystującego zaawansowane metody przetwarzania obrazów i uczenia maszynowego do określenia struktury tkankowej ludzkiego naskórka na zdjęciach histopatologicznych preparatów skóry barwionych hematoksyliną-eozyną oraz do opracowania systemu automatycznej diagnostyki wybranych typów melanocytowych zmian skórnych na podstawie analizy morfometrii naskórka w danej zmianie. Według najlepszej wiedzy Autora to pierwsza próba automatycznego wyznaczania parametrów morfometrycznych naskórka i wykorzystania tej wiedzy w celach diagnostycznych.

W rozprawie postawiono następującą tezę główną oraz tezy pomocnicze:

Teza główna Zaawansowane metody przetwarzania obrazów oraz uczenia maszynowego umożliwiają szczegółową analizę morfometrii naskórka zmiany skórnej na obrazach histopatologicznych.

Teza I Przy wykorzystaniu zaproponowanych metod segmentacji i analizy obrazu możliwa jest ocena stopnia złośliwości zmiany skórnej.

Teza II Metody uczenia maszynowego pozwalają na sformułowanie podstawowych kryteriów diagnostycznych, możliwych do zastosowania podczas badań histopatologicznych.

Zaproponowane rozwiązanie

Zaproponowane metody oraz cały system diagnostyczny zostały zaprojektowane w celu automatycznej analizy i rozróżniania trzech podstawowych typów skórnych zmian melanocytowych: plam soczewicowatych (zmiany łagodne), znamion łączących dysplastycznych (formy pomiędzy zmianami łagodnymi a złośliwymi) oraz czerniaków (zmiany złośliwe).

Ze względu na znaczący wzrost na całym świecie wskaźników zachorowań na czerniaka skóry, najgroźniejszej odmiany raka skóry, i spowodowanych nim zgonów wśród populacji rasy białej ewidentną staje się potrzeba opracowania skutecznych i wydajnych (automatycznych) narzędzi diagnostycznych dla dermatopatologii. Tradycyjny, ręczny sposób diagnozowania zmian skórnych – badanie histopatologiczne – ma trzy główne wady: jest podatny na tzw. błąd ludzki, jego wyniki często nie są powtarzalne, a analiza dużych partii preparatów to żmudna czynność. Dzięki zastosowaniu metod wizji komputerowej i uczenia maszynowego do analizy cyfrowych zdjęć preparatów o dużej rozdzielczości możliwe jest rozwiązanie wszystkich trzech wspomnianych problemów – poprawa skuteczności diag-

nostycznej, zwiększenie powtarzalności uzyskiwanych wyników diagnostycznych, oraz skrócenie czasu niezbędnego na postawienie diagnozy.

W ramach prowadzonych badań opracowano, zaimplementowano i przetestowano poniższe algorytmy oparte na metodach przetwarzania obrazów:

- Automatyczna segmentacja obszaru preparatu histopatologicznego (tkanki) oparta o analizę statystyczną rozkładu koloru w przestrzeni barw CIELAB z wykorzystaniem operacji morfologicznych. Analizowany jest zarówno cały obraz, jak i lokalne otoczenie jego wybranych fragmentów.
- Automatyczna metoda segmentacji warstwy naskórka oparta na analizie rozmieszczenia i kształtów obszarów tła obrazu w obrębie tkanki oraz histogramu stężeń hematoksyliny i eozyny w preparacie, wykorzystująca metody statystyczne i przetwarzania obrazów.
- Automatyczna metoda segmentacji sopli naskórkowych w warstwie naskórka oparta przede wszystkim na analizie krzywizny konturu obszaru celem wykrycia potencjalnych krańców podstaw wypustek oraz na analizie szkieletu obszaru w reprezentacji grafowej (z wykorzystaniem algorytmów grafowych i operacji morfologicznych) celem ostatecznego wyznaczenia podstaw i ich położenia względem osi głównej obiektu.

Każdy z powyższych algorytmów jest w pełni zautomatyzowany i w dużym stopniu oparty na wiedzy dziedzinowej z zakresu morfometrii i właściwości biochemicznych skórnych struktur tkankowych.

Następnie opracowano parametry opisujące morfometrię całej warstwy naskórka w oparciu o umiejscowienie wzdłuż osi głównej warstwy i wartości podstawowych parametrów morfometrycznych sopli naskórkowych. Pozwoliło to na zaproponowanie algorytmu automatycznej klasyfikacji zmian skórnych na podstawie morfometrii ich regionu naskórka, z wykorzystaniem metod uczenia maszynowego – w pracy porównano skuteczność czterech wybranych klasyfikatorów, przy czym najskuteczniejszym okazał się model maszyny wektorów wspierających (SVM). Dokonano również analizy statystycznej zależności między morfometrią naskórka a podstawowymi klasami melanocytowych zmian skórnych z użyciem modelu drzewa decyzyjnego, w celu określenia kryteriów diagnostycznych pod kątem badania histopatologicznego.

Skuteczność systemu została zweryfikowana z użyciem bazy składającej się z 75 zdjęć całych preparatów skórnych zmian melanocytowych o wysokiej rozdzielczości (po 25 dla każdego typu zmiany), opisanych przez doświadczonego dermatopatologa. System uzyskał dokładność 74.7% i 86.7% odpowiednio dla zadań klasyfikacji „plama soczewicowata, znamię łączące dysplastyczne, czy czerniak” i „zmiana łagodna czy złośliwa”. Dla tych samych zadań klasyfikacji model drzewa decyzyjnego (wyuczony z użyciem wartości zaproponowanych wskaźników morfometrycznych) uzyskał dokładność 73.3% i 88.0%, i tym samym może zostać użyty do sformułowania kryteriów diagnostycznych związanych z morfometrią naskórka.

Zdaniem Autora uzyskane wyniki uprawniają go do uznania, że tezy postawione w rozprawie zostały dowiedzione.

Table of Contents

List of Symbols	14
1. Introduction	17
1.1. Research objectives	19
1.2. Pioneering nature of the study	20
1.3. The general workflow of the proposed diagnostic algorithm	21
1.4. Structure of the dissertation	23
1.5. Research projects related to the study	23
1.6. Language used in the dissertation	24
2. Medical background	25
2.1. Histology and histopathology	25
2.2. Preparation of slides	26
2.2.1. Key stages in preparing histology slides	26
2.2.2. Hematoxylin and eosin (H&E) staining	27
2.2.3. Quality control in routine H&E staining	27
2.2.4. Slide artifacts	28
2.3. Digital pathology	31
2.3.1. Whole slide imaging (WSI)	31
2.3.2. Imaging devices	32
2.3.3. Reliability of WSI systems (in comparison to light microscopy)	34
2.3.4. Stain color deconvolution and normalization	35
2.4. Skin histology	37
2.4.1. Epidermis	38
2.4.2. Dermis	40
2.4.3. Hypodermis	42
2.4.4. Appendages	43
2.4.5. Morphometry of cells and tissue structures	44
2.5. Skin pathology	45

2.6.	Skin melanocytic lesions	46
2.6.1.	Benign lesions: Lentigo maligna.....	47
2.6.2.	Dysplastic lesions: Junctional dysplastic nevus.....	47
2.6.3.	Malignant lesions: Melanomas	49
2.6.4.	Differential diagnosis between benign and malignant lesions.....	50
2.7.	Melanoma misdiagnosis problem.....	52
3.	Preface to the proposed algorithms	55
3.1.	Related works in the field	56
3.1.1.	Tissue segmentation	56
3.1.2.	Epidermis segmentation	56
3.1.3.	Epidermis morphometry	58
3.2.	Datasets specification.....	59
4.	Tissue segmentation	61
4.1.	Algorithm for tissue segmentation.....	62
4.1.1.	Preprocessing	62
4.1.2.	CIELAB colorspace metrics	62
4.1.3.	Determining the “mean” background color	63
4.1.4.	Initial background segmentation.....	64
4.1.5.	Refinement of region boundaries	65
4.2.	Results.....	66
4.2.1.	Statistical analysis	66
4.2.2.	Parameter selection	67
4.2.3.	Experimental results.....	68
4.3.	Conclusions and discussion	71
4.4.	Summary.....	72
5.	Epidermis segmentation	73
5.1.	Algorithm for automatic epidermis segmentation	73
5.1.1.	Specimen segmentation.....	74
5.1.2.	Porosity analysis	74
5.1.3.	Stain concentrations analysis	78
5.1.4.	Final refinement	84
5.2.	Results.....	85
5.2.1.	Parameters selection.....	85
5.2.2.	Evaluation metrics.....	85

5.2.3. Segmentation performance.....	86
5.3. Conclusions and discussion	87
5.4. Summary	89
6. Epidermis morphometry	91
6.1. Algorithm for retes segmentation and for their morphometry measurement	91
6.1.1. Determining the epidermis “outer” edge	92
6.1.2. Computing the epidermis skeleton.....	94
6.1.3. Determining the epidermis main axis	95
6.1.4. Retes segmentation	97
6.1.5. Retouching the main axis	106
6.1.6. Computing the morphometry of retes	107
6.2. Results.....	108
6.2.1. Evaluation metrics.....	108
6.2.2. Performance	109
6.3. Conclusions and discussion	112
6.4. Summary	114
7. Feature extraction and classification.....	115
7.1. Characteristic features for epidermal morphometry	115
7.2. Classification.....	117
7.2.1. Multinomial logistic regression	118
7.2.2. Naïve Bayes classifier	118
7.2.3. Classification tree	119
7.2.4. Support vector machine.....	121
7.3. Results and discussion	123
8. Summary	129
8.1. Original contribution.....	129
8.2. Conclusions.....	130
8.3. Future research directions.....	132
Bibliography	133

List of Symbols

Image

I – 2D color image

R_I – image resolution

P – pixel (a point in an image)

c – color (represented as a vector of color components)

$I_1 \cdot I_2$ – (element-wise) multiplication of the image I_1 with the image I_2

I_1 / I_2 – (element-wise) division of the image I_1 by the image I_2

$I_1 * I_2$ – convolution of images I_1 and I_2

Morphometry

r – radius

L – length (e.g., of a path)

d – thickness, distance

A – area

$A(O)$ – area of the object O

Binary morphology

M – binary mask

$\overline{M_x}$ – negation of the mask M_x

$|M_x|$ – number of 1s in the M_x mask

SE – structuring element (of any shape)

$SE_D(r)$ – disk-shaped structuring element of radius r

$SE_{Sq}(a)$ – square-shaped structuring element of size $a \times a$

$M_x \oplus SE$ – dilation of the mask M_x using a structuring element SE

$M_x \circ SE$ – opening of the mask M_x using a structuring element SE

$M_x \bullet SE$ – closing of the mask M_x using a structuring element SE

$M_{\text{marker}} \xrightarrow{n} M_{\text{mask}}$ – reconstruction of the mask M_{marker} under the M_{mask} mask
(with n -connected neighborhood)

1. Introduction

Methods of computer vision and machine learning allow not only for automation of different tasks in various fields, but also for obtaining unbiased and reproducible results. One of such areas of application for advanced methods of computer vision and machine learning is a rapidly growing field of pathology called “digital pathology”. Pathology is the science of the causes and effects of diseases or injuries. When used in the context of modern medical treatment, the term typically refers to processes and tests addressing four components of disease: cause, mechanisms of development, structural alterations of cells, and the consequences of changes (its clinical manifestations) [1]. Traditionally, pathologists examined histopathological slides personally (and, typically, individually), under a light microscope. Such an approach has three main drawbacks: the quality of the diagnosis depends on the knowledge and skills of a particular pathologist, the results of a histopathological examination are often non-reproducible (especially when it comes to such tasks, as counting individual cells in a large tissue specimen), and quantitative analysis of a large volume of specimen is laborious [2, 3]. Digital pathology is mainly about capturing and analyzing whole slide images, i.e., glass slides of tissue specimens digitized at high magnification, and is driven primarily by developments in technologies. By applying the methods of computer vision and machine learning to digitized slides it is possible to address these issues – to improve diagnostic accuracy, to increase the reproducibility of diagnostic outcomes, as well as to shorten the time needed for making a diagnosis [4].

The need for more robust and more efficient (automatic) diagnostic tools is evident particularly in dermatopathology, where they could be used to diagnose skin cancers. Over the past several decades, there has been a significant and constant increase in the incidence rate of and mortality from cutaneous melanoma, the most aggressive and dangerous skin cancer, among Caucasian populations worldwide (Fig. 1.1) [5]. Despite that only roughly 4% of all diagnosed skin cancers are melanomas, melanoma is responsible for up to 70–75% of skin cancer-related deaths in the United States and Australia [6, 7]. Also in Poland, within the last two decades (1999–2016), this trend has been particularly alarming among the part of population in the economically productive age: melanoma incidence rate both among young people (20–44 y.o.) and among middle-aged people (45–64 y.o.) doubled [8]. In 2016, the melanoma death rate in Poland amounted to 1270, constituting about 1.3% of all cancer deaths [8]. Most of the increase in melanoma incidence rate in recent years is related to cases of thin superficial spreading melanoma (i.e., early melanomas), which constitute up to 70% of all diagnosed skin melanoma cases [9].

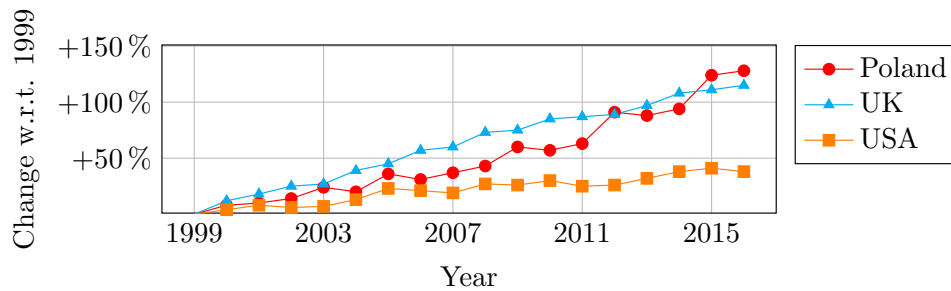


Figure 1.1. The increase in incidence rate of skin melanoma observed in UK, Poland and USA between 1999 and 2013 [8, 10, 11]. The incidence rate reported in 1999 has been set as a point of reference.

Two of the most important factors considered to result in melanoma are the brief but intense sun exposure pattern and unhealthy living habits. Due to a constant depletion of ozone layer in stratosphere resulting in higher exposure to UV radiation and due to unhealthy nutrition habits, in the future skin melanoma is likely to become one of the most common malignant neoplasms, with incidence rate even 2–10 times higher than nowadays [5, 12]. As no effective treatment of melanoma in an advanced stage has been developed so far, the early diagnosis of melanoma is of uttermost importance. When detected early, melanoma is treatable in nearly all cases with a simple surgical excision [12]. The progress in melanoma treatment is visible not only in the primary research but also in the development of sophisticated, more accurate methods of image processing, classification, and computer-aided diagnosis [13].

Nowadays, the histopathological examination constitutes the gold standard in skin melanoma diagnosis. Other forms of examination, such as dermatoscopy, are useful for screening, but their diagnostic confidence is inferior to the histopathological examination. Nevertheless, studies have shown that even this gold standard method has its serious drawbacks resulting in up to 25% of skin melanoma cases being misdiagnosed as benign lesions [2, 3, 14, 15]. The costs of such a misdiagnosis are significant, as melanoma tends to spread quickly beyond the local area into other parts of the body, including internal organs. A false-negative melanoma diagnosis also delays the beginning of treatment; when a subsequent examination is carried out the condition is often terminal. One of the reasons for such a high misdiagnosis rate is a vague definition of many routinely used histopathological criteria.

Pathologists, based on their personal clinical experience, proposed diagnostic criteria related to cellular atypia and morphology of lesions. In particular, they conjectured that there exist relationships between certain types of epidermal distortions and specific skin conditions (for instance, atrophy of basal epidermal unit is often sighted in melanomas, whereas uniform elongation of rete ridges tend to indicate a benign lesion) [16, 17]. In recent years numerous works in the field of medical image processing regarding automatic diagnosis of melanocytic nevi using clinical and dermoscopic images have been published [18, 19, 20] and the application of the deep learning approach significantly increased the accuracy of methods for automatic classification of skin melanocytic lesions [21, 22, 23]. **However, so far no thorough studies have been conducted neither on the actual level of connection between those symptoms and disease entities nor on the relation between the degree of distortion and the likelihood of given conditions [17].**

In the light of above data, both prevention and early diagnosis of melanoma became an extremely important issue. There is a high demand to develop computer-aided diagnostic systems facilitating the early detection of melanoma, which would lower its misdiagnosis rate.

1.1. Research objectives

The main aim of the proposed research project is to prove the major thesis:

By means of advanced image processing methods and machine learning algorithms it is possible to analyze the tissue structure of human epidermis in histopathological images.

To give proof of the major thesis the following minor theses have been proposed:

Thesis 1. Using the proposed methods for image segmentation and image analysis it is possible to assess the degree of malignancy of a given skin lesion.

Thesis 2. By means of machine learning methods it is possible to formulate the basic histopathological criteria for diagnosing skin lesions.

In order to prove it the following research steps have been carried out:

- I have performed the state-of-the-art review in order to provide a critical survey of the literature published in the past decade in the field of digital dermatopathology (in particular about epidermis segmentation), imaging equipment used in pathology, computer vision, image analysis and segmentation methods, as well as graphs.
- I have created a dataset consisting of 75 high-resolution whole slide images of skin melanocytic lesions together with the respective ground truth data relevant to the study (e.g., manually segmented slide background, epidermal regions, and rete bases).
- I have developed an automated method for tissue segmentation in high-resolution whole slide images of H&E-stained skin specimens.

The accurate tissue segmentation is usually an obligatory first step in computerized analysis of such images, e.g., to improve the accuracy of stain deconvolution methods or to perform morphological analysis of the tissue structure.

- I have developed an accurate automated method for segmentation of the epidermis in whole slide images of H&E-stained skin specimens.

The method permits to increase effectiveness of other diagnostics algorithms (e.g., those detecting and analyzing distribution of melanocytes in the epidermis).

- I have developed an automated method for measuring epidermal morphometry (its hypoplasia, hiperplasia, asymmetry; elongation and deformation of rete ridges), which helps to investigate geometric properties of epidermis in various skin conditions. In particular, the results of such a method help to organize information and knowledge on differentiating malignant from benign lesions.

- I have proposed a classifier which, based on morphometric parameters of the epidermis, could distinguish between three most common classes of melanocytic neoplasms (according to their metastatic potential): benign, dysplastic, and malignant.

In this dissertation the proposed and presented methods have been described, and furthermore their effectiveness has been evaluated and verified in experimental studies. Based on the obtained results the Author states that all the proposed theses have been proven.

To conclude, the results of this research could help to curb the influence of two main factors behind the high melanoma misdiagnosis rate: the histopathological similarity of less common melanoma subtypes to some benign nevi (leading to difficulties in diagnosis), and the fact that lesions are often evaluated not by dermatopathologist or surgical pathologist but by general histopathologists (who lack an in-depth knowledge of nuances between less common types of melanoma). The outcomes of the proposed project would form a basis for development of a system to diagnose specific types of skin conditions (e.g., specific types of melanoma), providing a solid ground for a treatment adjusted the exact condition. Moreover, the proposed solutions (e.g., for measuring epidermis morphometry) could be refined and used in different fields of computer vision, such as general object analysis and recognition.

1.2. Pioneering nature of the study

To the best of my knowledge, the conducted researches are pioneering ones regarding the following issues:

- The proposed method for tissue segmentation is the first one addressing this issue for high-resolution images. Unlike similar solutions for low-resolution images, the initial segmentation in the proposed approach is performed based not only local, but also global statistical information about color distribution in the image extracted from a 3D histogram of pixel intensities in the CIELAB color space. Consequently, the proposed method is more robust to variations in slide illumination.
- The proposed method for epidermis segmentation is the first approach directly addressing such issues as variations in slide illumination and staining. Unlike other methods described in the literature, which either analyze a single color channel or perform a fusion of information from several channels on a local (typically: per-pixel) basis, my method directly uses the whole available color information by the thorough analysis of the joint histogram of stain distributions (both hematoxylin and eosin). The proposed algorithm utilizes domain-specific knowledge of morphometric and biochemical properties of skin tissue elements to segment the relevant histopathological structures in human skin, what increases its robustness [24].
- The proposed method for automated measuring of epidermal morphometry is the first method addressing this issue. It makes an extensive use of graph algorithms applied on object's skeleton in order to analyze object's structure. Most of the proposed methodology is a general-purpose solution which can be applied to a wide range of problems in computer vision related to shape analysis.

- The study takes into account not only common benign and malignant neoplasms, but also the so-called “dysplastic” nevi which exhibit increased risk of turning into a malignant tumor than “normal” nevi. The existing studies are focused only on the dichotomic classification of a lesion as either benign or malignant.

1.3. The general workflow of the proposed diagnostic algorithm

The proposed automatic diagnostic algorithm consists of the following image processing and analysis stages (Fig. 1.2):

- **(Input).** The input to the algorithm is a high-resolution RGB whole slide image of a H&E-stained skin specimen.
- **Color thresholding and reconstruction in CIELAB color space for tissue segmentation.** The tissue regions are indirectly segmented by detecting the slide background regions using mainly color thresholding and reconstruction in CIELAB color space.

This stage uses a refined version of the technique I described in [25]: **P. Kłeczek**, J. Jaworek-Korjakowska, M. Gorgon. *A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images*. Computerized Medical Imaging and Graphics, 2020, vol. 79, 2022, Art. ID 101686, doi: 10.1016/j.compmedimag.2019.101686 [IF (2018) = 3.298]

- **Shape, distribution and color analysis for epidermis segmentation.** Epidermis is segmented based on information about shape and arrangement of slide background regions in an image and information about distribution and concentration of hematoxylin and eosin stains. Domain-specific knowledge of morphometric and biochemical properties of skin tissue elements is utilized to segment the relevant histopathological structures in human skin.

This stage uses a refined version of the technique I described in [26]: **P. Kłeczek**, G. Dyduch, J. Jaworek-Korjakowska, R. Tadeusiewicz. *Automated epidermis segmentation in histopathological images of human skin stained with hematoxylin and eosin*. Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology, 101400M (2017), doi:10.1117/12.2249018 [WoS]

and my ink segmentation method described in [27]: **P. Kłeczek**, M. Lech, G. Dyduch, J. Jaworek-Korjakowska, R. Tadeusiewicz. *Segmentation of black ink and melanin in skin histopathological images*. Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 105811A (2018), doi: 10.1117/12.2292859 [WoS]

- **Skeleton analysis for retes segmentation.** The basic morphometric features of rete ridges in a segmented epidermis (i.e., their location, base width, length, and height) are automatically determined based mainly on analyzing the curvature of the epidermal boundary to identify endpoints of projection bases, and on analyzing the skeleton of the epidermal region in the graph representation to identify rete bases and the location of retes along the epidermis main axis.

This stage uses a refined version of the technique I described in [24]: **P. Kleczek**, G. Dyduch, A. Graczyk-Jarzyńska, J. Jaworek-Korjakowska. *A New Approach to Border Irregularity Assessment with Application in Skin Pathology*. Applied Sciences (Basel), 2019, 9(10), 2022, doi:10.3390/app9102022 [IF (2018) = 2.217]

- **Feature extraction and classification.** Based on the morphometry of individual retes and their location along the lesion features describing the overall epidermal morphometry are proposed and computed. A machine learning classification algorithm uses the aforementioned features to make a diagnosis.
- **(Output).** The implemented computer-aided diagnostic system for the analysis of skin WSIs yields one of the following diagnoses: lentigo, dysplastic nevus, or melanoma.

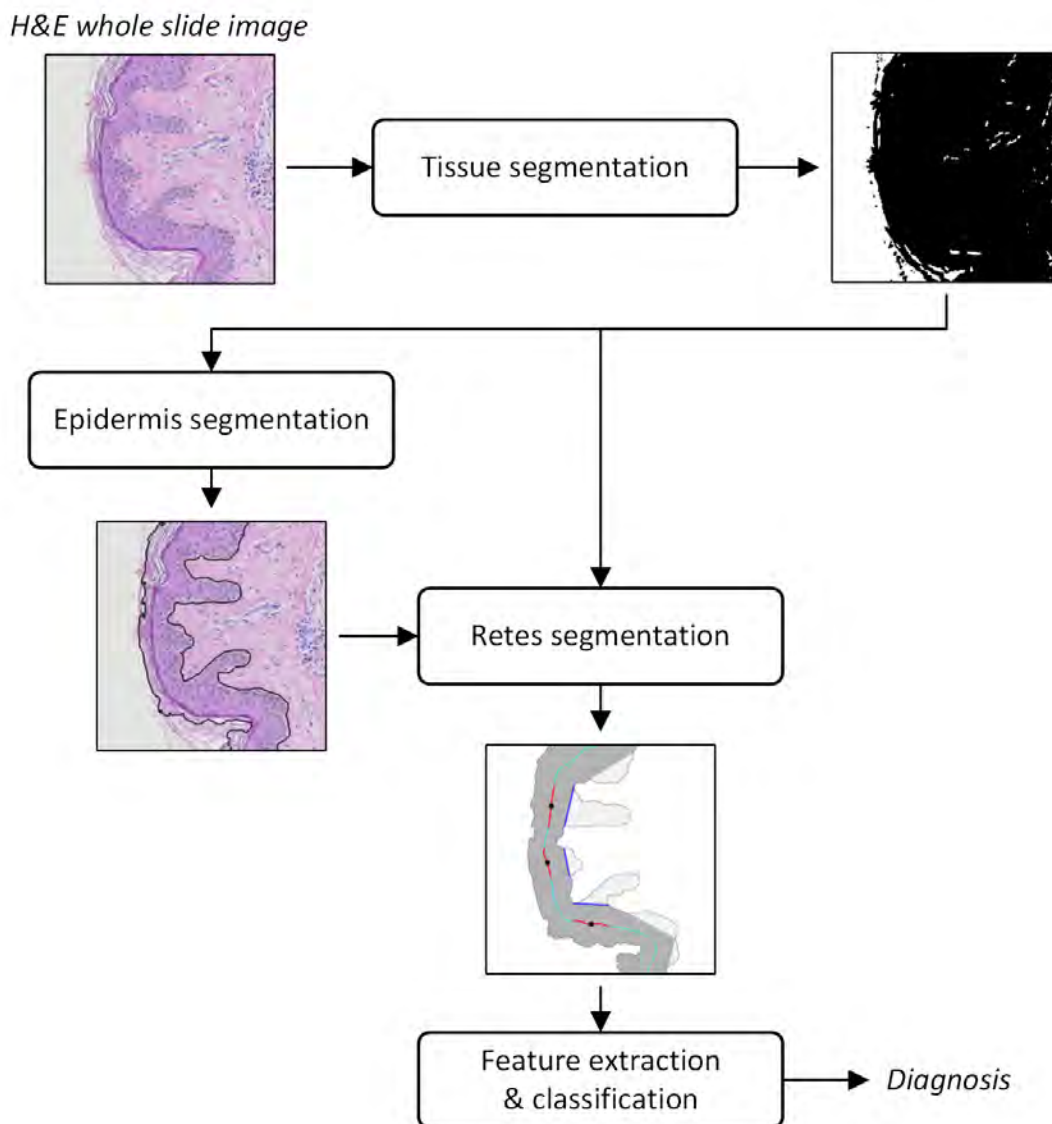


Figure 1.2. The general workflow of the proposed automatic diagnostic algorithm

1.4. Structure of the dissertation

The subsequent part of this dissertation is organized in six chapters as follows. Chapter 2 covers the basics of skin histology and dermatopathology, the preparation of histological slides and their digitization, the key histopathological diagnostic features of the selected skin melanocytic lesions, and the most important issues in diagnosing neoplasms. Starting from Chapter 3 the original algorithms have been described. Chapter 3 summarizes the state-of-the-art literature and challenges related to each stage of the proposed computer-aided diagnostic system, briefly outlines methods proposed by the Author for each stage, and characterizes datasets used in the study. Chapter 4 specifies tissue segmentation algorithm based on color statistical analysis, CIELAB color thresholding, and binary morphology. The results are then used to segment both epidermis and individual retes. In Chapter 5 an epidermis segmentation algorithm is presented, which is based on analyzing the information about shape and distribution of transparent regions in a slide image and information about distribution and concentration of hematoxylin and eosin stains. Chapter 6 covers the issue of segmenting individual retes and measuring their basic morphometry. In Chapter 7 the features describing the morphometry of the whole epidermal component have been proposed and performance of four machine learning classification algorithms used for lesion diagnostics have been compared. In particular, specific diagnostic guidelines for clinicians based on the output of the classification tree model have been presented. Chapter 8 closes the dissertation and highlights future research directions.

1.5. Research projects related to the study

The research tasks of the study were accomplished within the following research grants:

- Preludium 12, awarded by the National Science Centre (Poland)
 - Project title: “The analysis of morphometry of human epidermis in histopathological images”
 - Implementation period: 2017-10-05 – (now)
 - Implementing entity: AGH UST Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
 - Project no.: 2016/23/N/ST7/01361
- AGH University of Science and Technology research grant (2016)
 - Project title: “Working out an algorithm for segmenting epidermis in histopathological images of skin melanocytic lesions”
 - Implementation period: 2016
 - Implementing entity: AGH UST Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering
 - Project no.: 15.11.120.813

1.6. Language used in the dissertation

Throughout this dissertation, when describing the proposed algorithms, I used the first-person plural form (i.e., “we”) instead of the first-person singular form (i.e., “I”) – this decision was dictated by the custom present in the Anglo-American scientific writing. However, I would like to emphasize it, that I am the lead author of all algorithms proposed in this dissertation.

2. Medical background

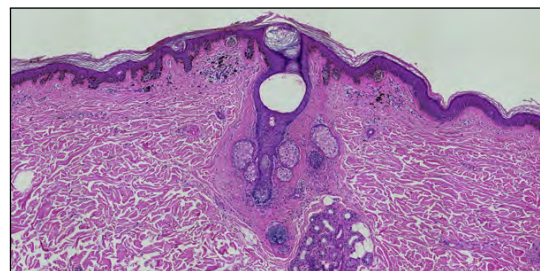
In this chapter we presented the basics of skin histology and dermatopathology, the process of preparing the histological slides (and its impact on tissues) and their digitization, as well as the key histopathological diagnostic features of the selected skin melanocytic lesions and the most important issues when classifying a neoplasm as either benign or malignant. All the presented information are relevant to the proposed automatic segmentation and classification methods, which heavily depend on the domain-specific knowledge.

2.1. Histology and histopathology

Histology, also called microscopic anatomy, is the scientific study of microscopic structures of tissues and organs of the body [28]. In order to distinguish different biological structures more easily and accurately, tissues are prepared using special processes called “histological techniques”; in particular, histological stains are often applied to add colors to or enhance the colors of certain types of biological structures (Fig. 2.1).



Clinical view



Histopathological image

Figure 2.1. Clinical view vs. histopathological image of a skin lesion

Pathology is the study of structural and functional abnormalities that are expressed as diseases of organs and systems [29]. Histopathology is the microscopic examination and study of diseased biological cells and tissues using the same set of techniques as in histology. Histopathology is used mainly in clinical medicine and it typically involves the examination of a biopsy (i.e., a surgically removed sample or specimen taken from a patient for the purposes of detailed study) by a specialist physician called a pathologist.

The histopathological examination is the gold standard for the diagnosis of skin diseases, especially for almost all types of cancer [30]. Other forms of examination, such as dermatoscopy and confocal

microscopy, are useful for screening, but their diagnostic confidence is inferior to a careful analysis of a histology slide. The histopathological examination of a skin lesion is carried out in particular always then when the result of a dermatoscopic examination suggest that the lesion might be a skin neoplasm [31].

2.2. Preparation of slides

This section discusses the key stages in preparing histology slides, the basics of hematoxylin and eosin (H&E) staining, the challenges in maintaining consistency of H&E staining, and the most common slide artifacts. All the presented information are related to the routine histopathological examination of skin specimens.

2.2.1. Key stages in preparing histology slides

The five main stages in the preparation of histology slides are fixing, processing, embedding, sectioning, and staining.

Fixing. Samples of biological tissue (Fig. 2.2a) are “fixed” to preserve tissue in a state as close to that of the living tissue as possible and to prevent postmortem decay (autolysis and putrefaction). This requires a chemical fixative that can stabilize proteins and nucleic acids of the tissue by making them insoluble. The fixation step for light microscopy is usually done by immersing the tissue in formalin for several hours to several days (Fig. 2.2b). Even when performed carefully, the fixation procedure always alters the sample to a certain extent potentially introducing artifacts that can hinder interpretation of images of fine cellular details.

Processing. Tissue processing is done to remove water from the tissues and replace it with a medium that solidifies, allowing extremely thin sections to be sliced. The medium of choice is wax. Firstly, the tissue is dehydrated with successively stronger concentrations of ethyl alcohol. Since alcohol does not mix with wax, in a subsequent clearing step alcohol in the tissue is replaced with a clearing agent (typically xylene), which will mix with wax. Finally, specimens are put into a warm paraffin wax bath for several hours, during which the paraffin enters each individual cell in the tissue.

Embedding. Before the tissues can be sliced they must be secured in a very hard solid block of wax (Fig. 2.2c).

Sectioning. To be able to clearly observe the detailed microstructure of the tissue using microscopy techniques, embedded tissue samples need to be sectioned into sufficiently thin slices. In case of samples to be studied using light microscopy, slices are typically 4–5 μm thick and are cut using a steel knife mounted in a microtome (Fig. 2.2d).

Staining. Finally, slices are treated with an appropriate histology stain, as all of the fixed materials have a similar refractive index and a similar, nearly transparent color when viewed using an ordinary optical microscope. Staining biological tissues is done to both increase the contrast of the tissue and also highlight some specific features of interest, depending on the type of tissue and the stain used.

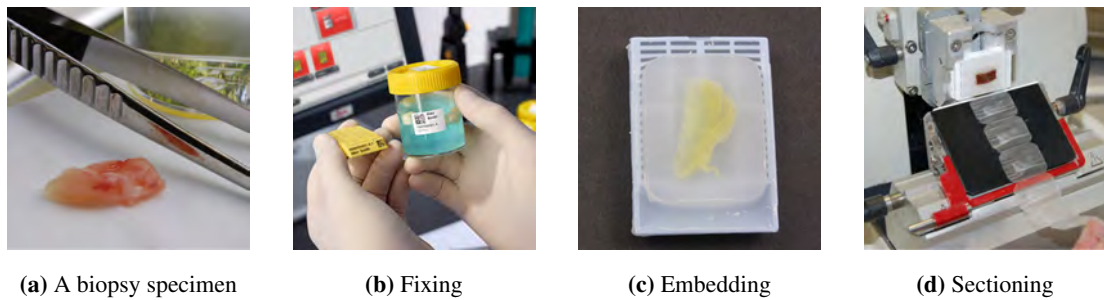


Figure 2.2. Selected stages of the slide preparation workflow [32]

2.2.2. Hematoxylin and eosin (H&E) staining

Hematoxylin and eosin stain (H&E stain) is the most commonly used stain for light microscopy in histopathology laboratories due to its comparative simplicity and ability to demonstrate a wide range of both normal and abnormal cell and tissue components [33].

The staining method involves application of hemalum, a complex formed from aluminium ions and hematein (oxidized hematoxylin), followed with a solution of eosin Y. The hematoxylin component binds to basophilic structures, such as DNA and RNA, and colors them blue (in skin it stains nuclei of cells, rough endoplasmic reticulum, ribosomes, keratohyalin granules, and elastic fibers) [34]. The eosin colors cell acidophilic structures, generally composed of intracellular or extracellular protein (e.g., cytoplasm and most connective tissue fibers, like collagen), in varying shades and intensities of pink, orange, and red [33]. Consequently, all relevant tissue structures in skin specimen are stained and effectively the whole area of the specimen appears in color, as shown in Figure 2.3 (some intrinsic pigments such as melanin absorbs neither hematoxylin nor eosin, yet they appear e.g., yellow and brown).

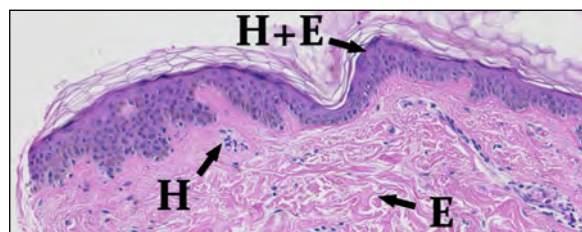


Figure 2.3. A skin section stained with hematoxylin (H) and eosin (E). In some tissue regions we observe the interaction between the two stains (H+E).

2.2.3. Quality control in routine H&E staining

In order to avoid difficult pathological interpretation and to be able to develop effective methods for automatic analysis of images of stained microscope slides, it is vital to maintain consistency of staining. Some of the most typical sources of variations in staining quality are discussed in the following paragraphs.

Times of staining procedures. The times for hematoxylin staining and for satisfactory differentiation will vary according to the type and age of alum hematoxylin used, the type of tissue, and the personal preference of the pathologist [33].

Type of hematoxylin. Hematoxylin can be prepared in numerous ways. For routine H&E staining of tissues, the most commonly used hematoxylin are Ehrlich's, Mayer's, Harris's, Gill's, Cole's, and Delafield's [33]. Figure 2.4 shows examples of staining with various hematoxylin.

Age of the stain. As with most of the chemically ripened alum hematoxylin, the quality of the nuclear staining begins to deteriorate after a few months [33].

Mode of hematoxylin staining. The alum hematoxylin can be used in one of two ways [33]:

- progressively, meaning that the section is stained with hematoxylin for a predetermined time to stain the nuclei adequately but leave the background tissue relatively unstained (e.g., Mayer's, Gill's I & II hematoxylin)
- regressively, meaning that the section is over-stained and then differentiated in acid alcohol, followed by 'blueing' (e.g., Harris hematoxylin)

Other factors. Other common factors introducing the variability in the staining quality include: a change of hematoxylin supplier, pH differences, fixation, variations in processing schedules, section thickness, and excessive hot-plate temperatures [33].

Some of the above-mentioned sources of variations may be eradicated or their influence minimized by using commercially prepared hematoxylin and eosin solutions as well as automated staining machines [33]. Automated staining machines allow accurate and consistent staining, differentiation, and dehydration by adjusting the times of each step.

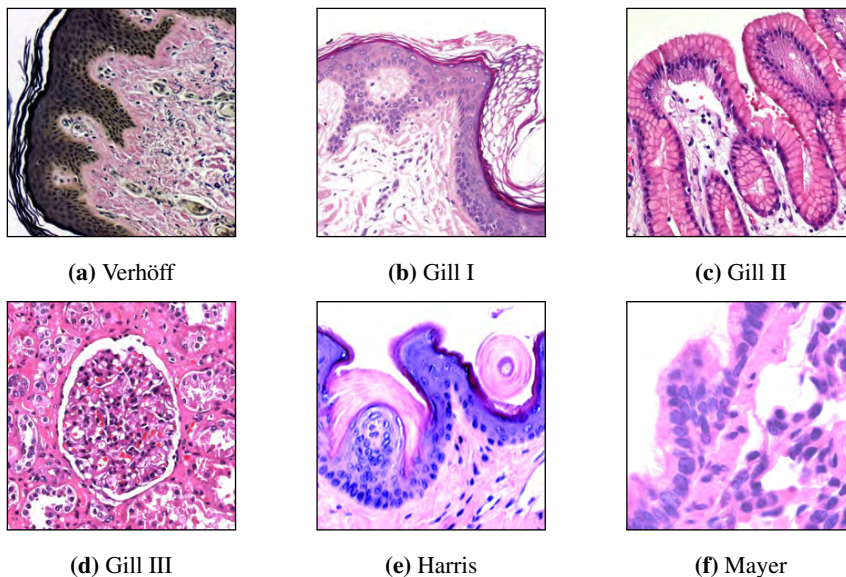


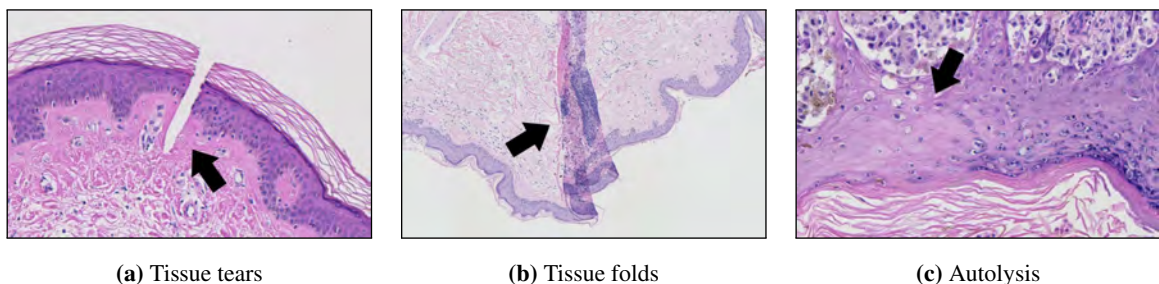
Figure 2.4. Staining with H&E using various hematoxylin.

2.2.4. Slide artifacts

The accurate diagnosis of lesions by microscopy requires such preparation of tissue sections, usually stained, that they represent their structures in life as closely as possible. However, usually pathologists encounter slides either improperly fixed or mishandled during tissue processing, resulting in alterations in tissue details. Such changes, referred to as “artifacts”, are the major source of diagnostic problems.

An artifact is an artificial structure or tissue alteration on a prepared microscopic slide caused by an extraneous factor, such as clinical application of chemicals, surgical mishandling of specimen, inadequate microtomy, inadequate tissue fixation or improper fixation medium, faulty tissue processing, or improper staining [35, 36]. Some of the most common artifacts are listed below.

Prefixation artifacts. *Squeeze artifact* is the form of tissue distortion resulting from the compression of tissue by forceps or other surgical instrument, and includes crush, hemorrhage, splits, fragmentation, and pseudo-cysts (Figures 2.5a and 2.5b) [36]. *Autolysis artifact* results from the delay in fixation or inadequate fixation, which leads to decay of tissue structures and produces changes like increased eosinophilia, shrunken appearance of cells (with cytoplasmic clumping), and impression of scar formation (Fig. 2.5c) [37].



(a) Tissue tears (b) Tissue folds (c) Autolysis

Figure 2.5. Prefixation artifacts: (a) tissue tear due to rough handling by forcep, (b) folds due to rough handling by forcep, and (c) tissue autolysis due to delayed fixation.

Tissue-processing artifacts. *Improper dehydration* can cause artifacts for one of the following reasons: too long treatment in higher concentration of alcohol results in high degree of shrinkage of the tissue due to overdehydration, whereas too long treatment in lower concentration of alcohol macerates the tissue and is seen as vacuolization [38]. These two procedures will also interfere with staining properties [39]. *Improper embedding* may give rise to cracks during sectioning for instance when the specimen is exposed for too long during embedding procedure, which causes excessive hardening so that the specimen becomes brittle [39]. Tissue-processing artifacts may occur even when slides are prepared by experienced technicians.

Artifacts related to microtomy. *Scores* and *tearing* in sections are caused either by a damaged knife edge or when sectioning hard particles (e.g., foci of calcification); in the former case the tear usually extends across the whole section (Fig. 2.6a). *Compression artifacts* are usually caused by one of the following: a blunt knife, bevel of the knife too wide, or wax too soft (Fig. 2.6b). *Alternate thick and thin sections* are produced when the wax is too soft, a block or knife is loose, the knife clearance angle is too fine, the mechanism of a microtome is faulty, or when the rhythm of cutting is not steady and smooth (Fig. 2.6c). *Poor sectioning* is practically unavoidable, as since tissue may be sliced at any point, not all tissue structures are sliced exactly at their midpoint (e.g., the sliced nuclei rarely consist of two hemispheres, but rather of two spherical caps). *Tangential cut* is a related problem; in tangentially cut tissue the connective tissue cores may become entrapped within the epithelium, giving a false impression of invasive squamous cell carcinoma (Fig. 2.6d). This sort of problems is typically unavoidable due to the structural nature of the skin.

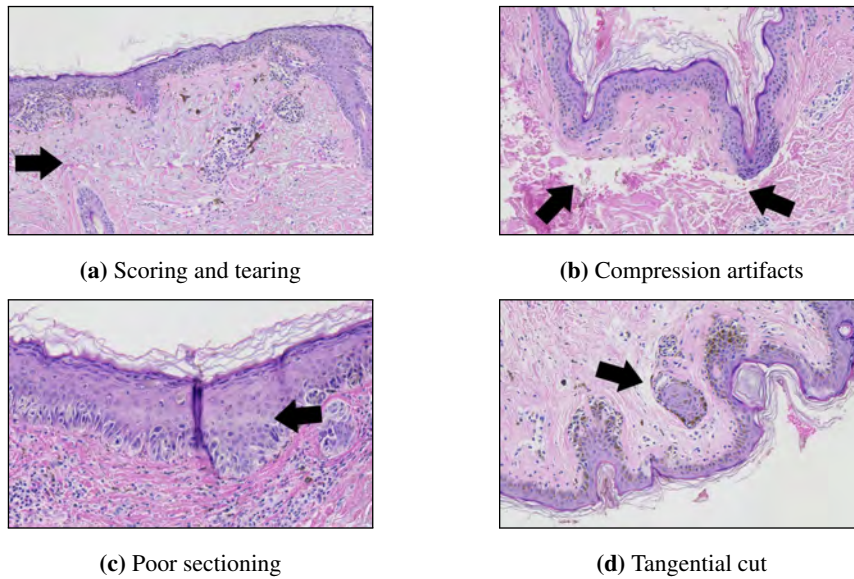


Figure 2.6. Artifacts related to microtomy: (a) scoring and tearing of section due to a nick in knife edge, (b) curling artifact due to folding of tissue due to blunt microtome knife, (c) a section showing thick and thin areas (thick areas stains more darkly than thin areas), (d) tangential cut artifacts.

Artifacts related to floatation and mounting. *Folds* and *wrinkles* in section occur when very thin paraffin sections are forced to stretch unevenly around other structures which have different consistencies; they appear as darker-stained strands (Fig. 2.7a). *Air bubbles* are formed under the cover-slip when the mounting medium is too thin (and, as it dries, air is pulled in under the coverslip) or when either clearing agents or coverslipping media are contaminated; areas where the cover-slip fails to adhere to the glass slide properly show slightly altered staining (Fig. 2.7b) [40].

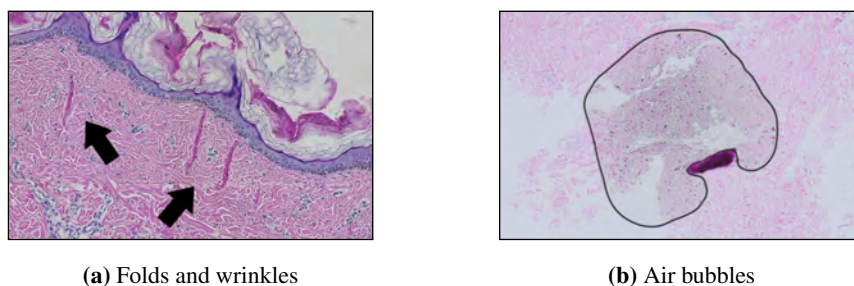


Figure 2.7. Artifacts related to floatation and mounting: (a) wrinkles and folds due to uneven stretching of tissue sections, (b) air bubbles formed during mounting procedure,

Staining artifacts. When the tissue has been fixed for a long time in a formalin fixative that has gradually become more acid or when excess amount of acetic acid is added (the effect of acetic acid is to increase ionization of tissue amino groups which results in more eosin attaching, which sometimes can improve the depth of coloration), it is problematic to get adequate nuclear staining with hematoxylin without also staining the cytoplasm; this phenomenon gives a uniformly muddy purple to the finished section after eosin has been applied [33].

2.3. Digital pathology

Although traditionally pathologists examined histopathological slides under a light microscope, quantitative analysis of a large volume of specimen is a laborious and difficult task. Moreover, since pathologists make the diagnosis based on their personal clinical experience, it is often subjective and leads to intra- and inter-observer variability [2, 3]. These issues may be addressed by digital pathology, a rapidly growing field primarily driven by developments in technology, offering such advantages as remote diagnostics and application of image analysis to improve efficiency and accuracy of the decision process [4]. Digital pathology is mainly about analyzing whole slide images (WSIs), i.e., glass slides of tissue specimens digitized at high magnification and thus able to provide global information for quantitative and qualitative image analysis.

The following sections discuss the idea of whole slide imaging, the devices needed to produce a whole slide image, the reliability of WSI systems (in comparison to “traditional” light microscopy), and the way to ensure consistent stain appearance among digital images of slides from different batches.

2.3.1. Whole slide imaging (WSI)

Whole slide imaging is the process in which entire histologic or cytologic slides are digitized at high resolution, typically using dedicated hardware (slide scanner). Such high-resolution whole slide images (WSIs) are able to provide global information for quantitative and qualitative image analysis [41]. In relation to WSI, virtual microscopy enables review of all regions of interest within the digitized slide at different magnifications using a personal computer with relevant software [42]. The whole process aims to emulate conventional light microscopy in a computer-generated manner. Figure 2.8 shows an example of a contemporary WSI scanner and virtual slide viewer, as well as an example of a WSI image.

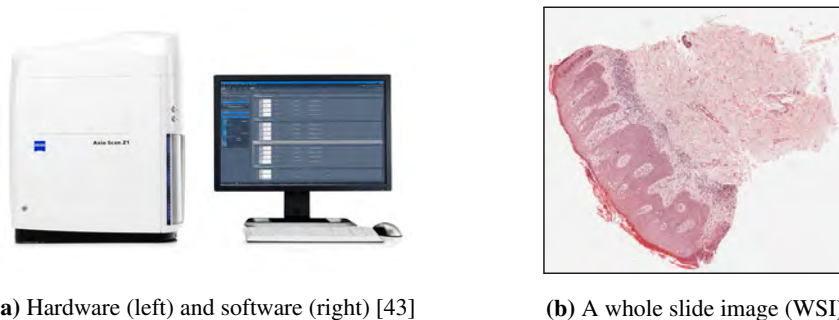


Figure 2.8. Whole slide imaging: (a) a WSI scanning system consists of a dedicated hardware and software; (b) an example of a WSI produced by a scanning system.

There are numerous advantages of making use of whole slide imaging, both in routine settings and in research image-related applications, for instance: WSI platforms have the potential to increase workflow efficiency and balance workloads [44], it is possible to integrate stored digital slides into patient’s electronic journal and to access them remotely [45], and digital slides can be analyzed using computer-aided diagnostic tools which can automate or quantify with greater consistency and accuracy than light microscopy [46]. Despite these benefits, the adoption of whole slide imaging by pathologists worldwide

has been slow mainly for such reasons as: the high cost of these systems and digital slide storage, their inability to handle high-throughput routine work, regulatory barriers in certain countries, and pathologists' reluctance to such systems [47].

Whole slide imaging is a relatively new technique, as the first automated, high-resolution whole-slide imaging system was developed only in 1999 by Wetzel and Gilbertson. However, during the last decade, a wide range of commercially available WSI instruments have been developed. A list of common WSI systems and their respective vendors as well as a list of differences between selected WSI systems can be found in [48].

For most diagnostic work the digital slides are routinely scanned at $\times 20$ magnification (at resolution approximately $0.25\text{--}0.5\ \mu\text{m}/\text{px}$), with sporadic use of $\times 40$ magnification for research purposes or for certain types of cases (e.g., for mitosis count in breast cancer), although $\times 40$ magnification is expected to soon become the standard [49, 50]. To enable optimized real-time viewing across various magnification (i.e., $\times 4$, $\times 20$, etc.) instantaneously, files are typically constructed into a multilayered "pyramid", which generates enormous image files [48].

Storing such high-resolution digital slides in an uncompressed format or using lossless compression can result in very large files (in the order of several GB), which impacts storage costs and work throughput, therefore for routine examination by pathologists a lossy compression technique is usually applied. For this, either the JPEG or JPEG 2000 image standards are most commonly used. Using JPEG compression with compression quality factor 70 results in images with acceptable file size compression artifacts unnoticeable for diagnostic purposes [50]. However, compressing images in a lossy way renders them practically useless for various methods of automatic digital image processing and analysis [51], therefore for this sort of research TIFF format is employed to archive virtual slides.

2.3.2. Imaging devices

In recent years rapid developments at all levels of microscopy (i.e., contrast, illumination, resolution, signal detection and data processing) have occurred and microscopes evolved into so-called imaging systems. Such systems are characterized mainly by a switch from analogue to digital cameras, by motorization of components, and advances in software-control of components and detectors/cameras. However, for almost all of today's imaging systems one aspect remained unchanged, the human factor. Limited human resources and human error pose problems not only in clinical and diagnostic workflows, but also in research environments, where reproducibility is critical. Automation for microscopic imaging helps to resolve these issues: it saves the researcher labor and time; it improves accuracy, quality and precision of imaging experiments; and last but not least it increases reproducibility.

The desirable automation features for advanced microscope systems, related to setup and acquisition steps, include: identification and calibration of the sample carrier; configuring channels/tracks according to dyes, fluorescent stainings or contrasts used; adjustment of additional imaging parameters (e.g., z-stack, laser-power, time intervals); spotting and marking the location(s)/object(s)/area(s) of interest in

the sample; finding and keeping the focus automatically; and stopping acquisition when imaging does not yield predefined quality criteria.

Modern microscopes have a wide variety of components that can be automated by means of electronic control: shutters, filter wheels, stages, light sources, and focus control. Moreover, it is essential that all these motorized components, sensors and input devices are properly integrated into a software environment. Finally, typical imaging systems provide means for advanced data processing and storage to further boost throughput and efficiency of such systems. Although it is possible to assemble an imaging system from scratch or to retrofit an existing microscope setup with these components, it is then problematic to ensure proper integration of hardware and software components purchased from different sources into a well-coordinated, efficient system. A more robust solution purchase a whole system from one of the major microscope manufacturers to secure the full integration of components which will perform superbly with proprietary software (Fig. 2.9a).

The “traditional” microscope systems has at least two serious limitations in the context of digital pathology: they allow to scan only individual slides (to scan another slide a manual change is required, which hinders work throughput), but more importantly, due to the open construction they do not provide a consistent imaging conditions (as ambient light in a laboratory room interferes with the light source of a microscope). These shortcomings were eliminated with the introduction of whole slide scanners.

A whole slide scanner is a specially designed microscope under robotic and computer control, with a highly specialized camera(s) containing advanced optical sensors and allowing for batch scanning of slides, which has all components assembled in a special casing (Fig. 2.9b). Therefore it can be considered an upgraded version of a microscope imaging system. All current WSI scanners consist of the following essential components: (1) a microscope with lens objectives, (2) illumination systems (light source), (3) robotics to load and move glass slides around, (4) digital camera(s) for image acquisition, (5) a focusing system that precisely places an image on a camera, (6) a computer, and (7) software to manipulate, manage, and view digital slides [42, 48].



(a) A motorized microscope [52]



(b) A whole slide scanner [43]

Figure 2.9. WSI imaging devices

After digital data are captured via the camera’s charge-coupled device (CCD), the virtual slide is assembled together from large numbers of image frames in one of the following ways, depending on the particular scanner being used: tiling, line scanning, dual sensor scanning, dynamic focusing, or array scanning (the process is performed automatically by a specialized imaging software). The most com-

monly used methods for accomplishing this operation are tile-based and line-based scanning (Fig. 2.10), whereas in “traditional” microscope systems only tile-based scanning is available.

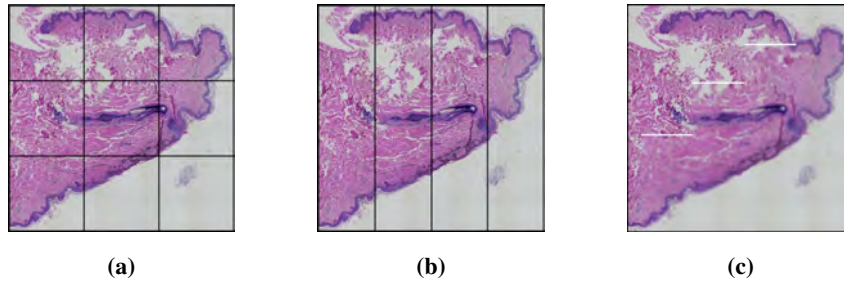


Figure 2.10. Scanning methods: (a) tile-based and (b) line-based; (c) line-based scanning of an actual glass slide shown while in progress.

Tile-based scanning relies on a robotics-controlled motorized slide stage to obtain square image frames in a checker pattern, with a slight overlap in each tile. Then, after the CCD-captured frames are “auto-correlated” with each other to ensure proper alignment, they are “stitched” together in a single, massive seamless image [48]. Line-based scanning utilizes a servomotor-based slide stage which moves sequentially in a linear fashion along a single axis of acquisition, producing a group of images in the form of long, uninterrupted strips. These strips are then aligned and “stitched” together into a single massive image. This method of slide acquisition greatly simplifies the image alignment process, since the number of image chunks and degrees of freedom associated with each is significantly reduced [48].

The output image resolution is determined by the magnitude of the microscope objective used for scanning (e.g., $\times 20$, $\times 40$), the numerical aperture of the objective, and the quality of the camera’s photosensors.

The scanning speed varies from 1–3 min per slide, depending on the objective magnification, and the number of z-stacks (plane of focus) acquired [48]. To speed up the scanning process, some devices allow for dynamic pre-focusing, which utilizes one camera to focus and another for scanning. Usually scanners offers a possibility to scan either an entire glass slide or, if desired, only a pre-selected region of interest on the slide.

Despite the evident advantages of whole slide scanners over “traditional” microscope systems, due to high costs of the former ones “traditional” microscope systems are still used in many laboratories, especially for research purposes.

2.3.3. Reliability of WSI systems (in comparison to light microscopy)

Many pathologists believe that the performance of digital pathology systems is inferior to that of light microscopy. However, numerous validation studies assessing intra- and inter-pathologist diagnostic discrepancy rates and diagnostic certainty between glass and digital slides (summarized in Table 2.1) concluded that there are no significant differences (with clinical or prognostic implications) in the diagnostic ability of the participants between the two modalities, and thus suggest that virtual slides are equivalent to conventional slides when making diagnoses in pathology. When discrepancies between WSI and light-microscopy diagnoses arise, they typically involve entities that are known to be subject of inter-observer

Table 2.1. The validation studies assessing intra- and inter-pathologist diagnostic discrepancy rates and diagnostic certainty between glass and digital slides by having a group pathologists review sets of cases by both light microscopy and digital pathology systems (digital slides were scanned at $\times 20$ magnification). The concordance rate include cases with full concordance, minor discordance, and discordance with no clinical significance.

Study	# of cases	# pathologists	Concordance [%]
Jukic et al. [54]	101	3	95.6
Gilbertson et al. [55]	25	3	100.0
Al-Janabi et al. [56]	100	6 ^a	94.0
Nielsen et al. [42]	96	4	96.4
Thorstenson et al. [49]	606	4	99.6

^a Every pathologist assessed his own cases.

variability [53]. Some mismatched diagnoses could be explained by the lack of the possibility to review the digital slides in a higher magnification than $\times 20$. A detailed review of the above-mentioned studies can be found in [53].

2.3.4. Stain color deconvolution and normalization

Variations in slide color are caused by a mosaic of section fixation procedures and staining protocols, diversity of types of registration equipment and their configuration, and more (as mentioned in Section 2.2). In particular, the H&E stain, the principal stain for examination in pathology (and particularly in dermatopathology), is hard to standardize and slides often contains preparation artifacts. Figure 2.11 shows examples of WSIs of H&E-stained skin specimens exhibiting huge differences in staining quality and background color (due to discrepancies in parameters of the light source in each of imaging devices). Therefore, stain color normalization is an important preprocessing step when developing methods for automatic analysis of histopathological images.

Stain deconvolution is the process of transforming a stained tissue section image from the normal RGB colour space into a series of stain channels, where each stain channel is a grayscale image representing the intensity of a particular stain expression across the original image, as shown in Figure 2.12 [57].

Standard histological stains absorb light and the amount of light absorbed in each spectrum is non-uniform, being a characteristic feature of a particular stain. Therefore, local stain concentrations and stain colors determine the appearance of an illuminated slide sample under the microscope. Areas where both no stain is present and the underlying tissue does not absorb a significant amount of light appear bright white, as all of the light will pass through. The proportions of each wavelength absorbed (i.e., absorption coefficients) form the stain vector of dimension equal to the number of wavelengths in the sensor used for imaging (three for a standard 24-bit RGB color camera). The stain vector varies significantly not

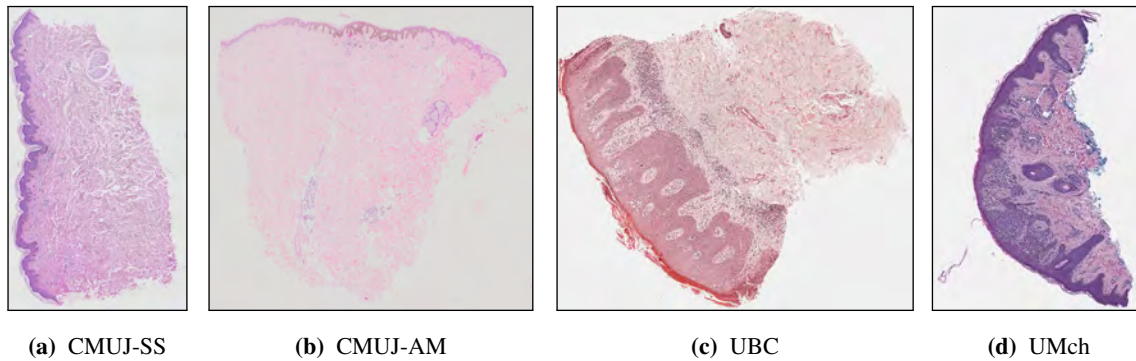


Figure 2.11. Examples of WSIs of H&E-stained skin specimens used in this study. Images (a), (c) and (d) were captured using different slide scanners, whereas image (b) was captured using an automatic microscope. Specimens origin from three different histopathology laboratories. Notice a huge difference in staining quality and background color (due to discrepancies in parameters of the light source in each of imaging devices) between those four sample images.

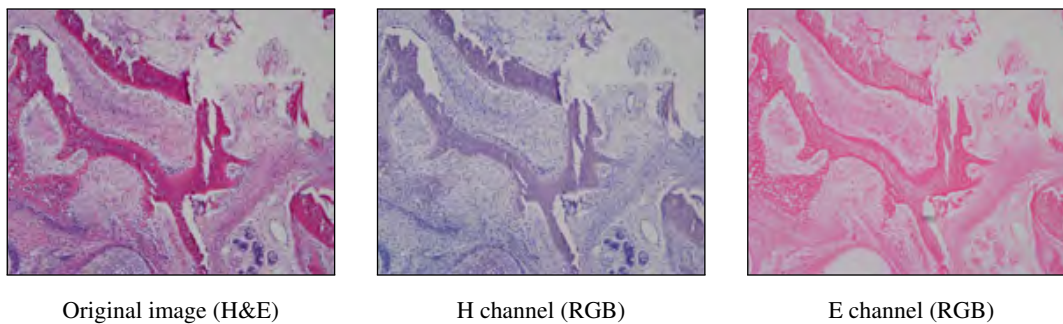


Figure 2.12. Stain deconvolution using the method proposed by Macenko et al. [58]. Grayscale stain channels were convolved with the respective stain vectors for readability.

only among different stains, but also for the same stain depending on numerous factors (as discussed in Section 2.2.3). A typical stain deconvolution method attempts to determine such an ideal stain matrix that its multiplication to the RGB color channels produces the desired stain channels. A stain matrix is composed of stain vectors, each vector representing the model color of a particular stain from the original image. In some applications, stain deconvolution is followed by a normalization step which aim is to standardize the stain color appearance across all the processed images before processing each stain color [59, 60].

The first stain deconvolution algorithm was proposed by Ruifrok and Johnston [61]. Their method outlined some key principles which are used even in more recent methods, like the use of stain matrices and the conversion of RGB color channels into optical density space. The authors also provided sample stain matrices for H&E and H&DAB, but these matrices were optimized for a particular set of images, whereas in general stain matrices need to be fine-tuned to the exact stain colors present in histopathological images (otherwise we may not achieve adequate deconvolution) [58].

Initial approaches to estimate image-specific stain matrices typically applied some statistical analysis to the color channels of an image and then reduced it into a series of stain vectors. Macenko et al. [58] proposed to estimate stain vectors by performing plane fitting in the optical density domain by means

of singular value decomposition (SVD) of the image data. Gavrilovic et al. [62] considered the problem on the Maxwellian chromaticity plane and assumed that projecting pixels onto this plane will result in perceptually similar colors appear close to each other (ideally, pixels corresponding to each stain are expected to appear in separable groups which could be modeled as a Gaussian mixture whose parameters are determined using an Expectation-Maximisation approach). Each stain vector is then estimated as the mean of its corresponding Gaussian distribution.

Some of more recent proposals take a supervised approach to stain deconvolution such as Khan et al. [63] and Alsubaie et al. [64]. These methods firstly use a pre-trained stain classifier to identify the locations where each stain is present, and then estimate stain vectors from these sets of classified pixels. The major drawback of supervised approaches is that to obtain satisfactory results a large set of annotated training data must be provided for a variety of stain types, which is often challenging to obtain.

Kather et al. [65] proposed to use principal component analysis (PCA) to get the optimal representation of stain colors by projecting the first two PCA components on the plane created by the stain vectors estimated using the pre-estimated stain matrix (provided by Ruifrok and Johnston [61]). However, PCA's assumption about the orthogonality between the main components may not be satisfied, especially in correlated stains such as hematoxylin and eosin, and using pre-estimated stain matrix usually results in suboptimal performance of stain deconvolution.

Recently, stain deconvolution algorithms based on independent component analysis (ICA) were also proposed, such as Trahearn et al. [66] and Alsubaie et al. [57]. These methods are based on the assumption that stain vectors can roughly be modeled as independent components according to the ICA model (the initial estimate is then corrected). Since ICA assumes that source signals are independent, which in case of histopathological images might not always be satisfied, Alsubaie et al. [57] relax this assumption by using ICA in the wavelet domain only on least Gaussian sub-bands.

Despite its deficiencies, the method proposed by Macenko et al. [58] remains one of the most commonly used approach to stain deconvolution, mainly due to its simplicity and an unsupervised approach [67]. Surveys on stain color deconvolution and normalization can be found in [57, 68, 69].

2.4. Skin histology

Skin is part of the integumentary system and is considered to be the largest organ of the human body. The three main layers of skin are: the epidermis, the dermis, and the hypodermis (Fig. 2.13a). The epidermis, a stratified squamous epithelium, is the outermost layer of skin. Deep to the epidermis lies the dermis, a thick layer of connective tissue consisting mainly of collagen and elastin, but also containing nerve endings, blood vessels, and appendages (such as hair shafts and glands). The deepest layer is hypodermis, consisting mainly of adipose tissue.

Each of the above-mentioned skin layers was characterized in the following sections based on information from [70, 71, 72, 73]. However, only details relevant to the topic of epidermis segmentation and

measuring epidermal morphometry were provided. In particular, components found only in glabrous skin (e.g., on palms, soles, and lips) were passed over.

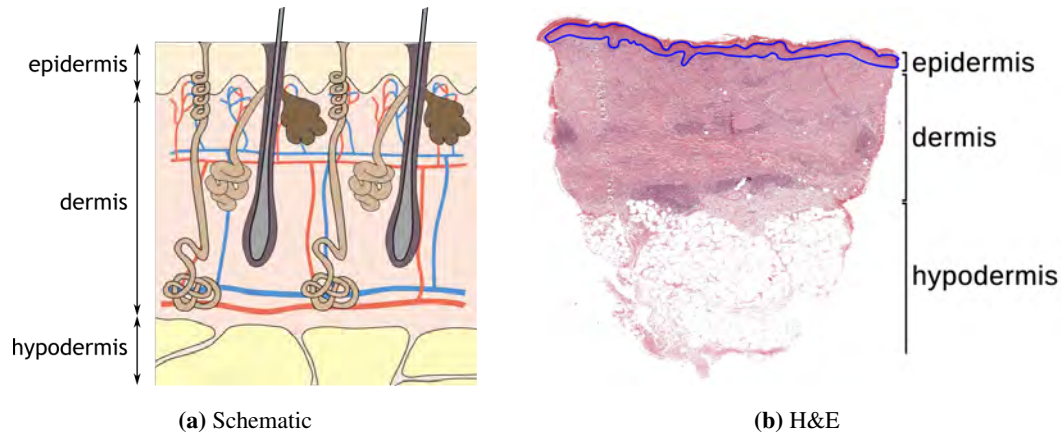


Figure 2.13. A skin cross-section. A whole slide image (WSI) of a typical skin preparation contains three main tissue layers: epidermis (marked blue), dermis and adipose tissue.

2.4.1. Epidermis

The epidermis is a stratified squamous epithelium, on average about 0.2 mm thick, composed of a variety of cell types – primarily of keratinocytes (90–95%), with minority populations of Langerhans cells, melanocytes, Merkel cells, and adendritic cells. It contains four to five layers, depending on its location (Fig. 2.14):

- **Stratum basalis** (basal cell layer): This layer is deepest and closest to the dermis. It contains melanocytes, a single row of keratinocytes (varying in shape from cubic to columnar), and stem cells, all resting on a basement membrane. The basement membrane serves as the point of demarcation between the epidermis and the dermis (i.e., the dermo-epidermal junction). Basal cells contain basophilic (or darkly staining) cytoplasm and an elliptical nucleus that is rich in chromatin.
- **Stratum spinosum** (prickle/suprabasal cell layer): This layer comprises most of the epidermis and contains approximately 4–10 layers of cells connected by desmosomes. Prickle cells are cuboidal in the lower layer and slightly flattened in the upper layers.
- **Stratum granulosum** (granular cell layer): This layer contains two or three layers of cells that contain basophilic lipid-rich granules. In this layer, cells begin to die and lose their nuclei, as they move away from the nutrients located in the deeper tissue. During the upward migration from the stratum spinosum, the cells of the stratum granulosum become flattened polygonal (even flatter than those in the suprabasal layer) and form two to three layers of pyknotic cells (where the nuclear material condenses).
- **Stratum lucidum**: This layer only exists in the thick skin of soles and palms and consists of mostly dead cells.
- **Stratum corneum** (keratin layer, horny cell layer): It is the most superficial layer of the skin, serving as a protective overcoat. The cells of this 10–20 layered stratum are dead, enucleated, flat, and filled

with aggregated keratin filaments. They are continuously lost from the skin surface and replaced by new cells emerging from the basal layer. Directly above the granular cell layer, the horny cell structure appears as an eosinophilic layer. The horny cell layer is very thick in the palms and soles.

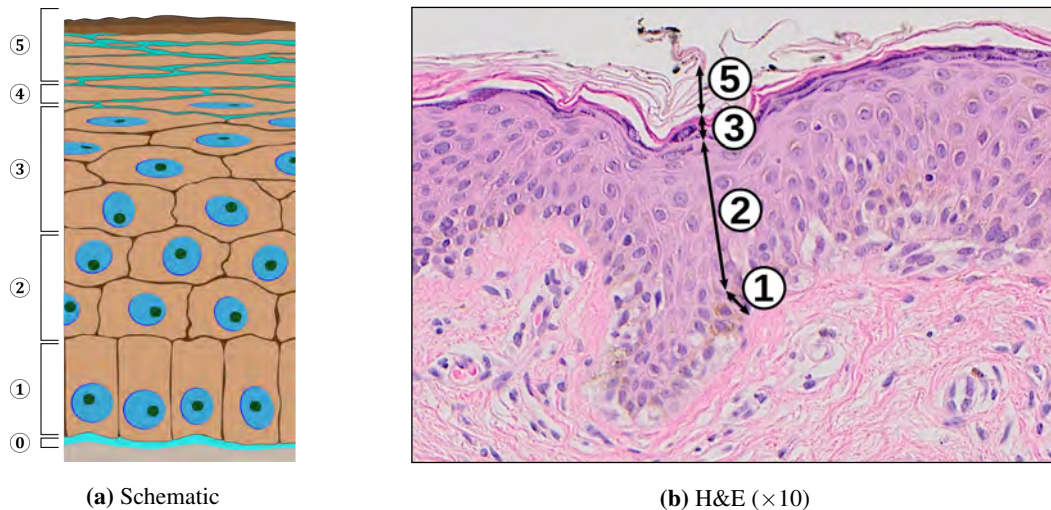


Figure 2.14. Structure of the epidermis: 0 – basement membrane, 1 – stratum basalis, 2 – stratum spinosum, 3 – stratum granulosum, 4 – stratum lucidum, 5 – stratum corneum.

A summary of microscopic appearance of cells forming the bulk of epidermis:

- **Keratinocytes** contain melanin granules that have been transferred to them by melanocytes (Fig. 2.15a).
- **Melanocytes** appear as clear cells in the basal layer with large, round, euchromatic nuclei. It is possible for a single melanocyte to provide melanin for approximately 30 keratinocytes via its dendritic arborisation (Fig. 2.15b).
- **Langerhans cells** are distributed throughout the strata spinosum and basale and can be readily identified by a hallmark tennis racket-like discoid Birbeck granule, in addition to its euchromatic cytoplasm (Fig. 2.15c).
- **Merkel cells** are clear, ovoid, and may occur singly or in clusters in the stratum basale of thick skin, on the palms of hands and soles of feet. They are difficult to tell apart from melanocytes.

A clear cell is any cell containing abundant glycogen or other material that is not stained by hematoxylin or eosin, so that the cell cytoplasm is very pale in routinely stained sections.

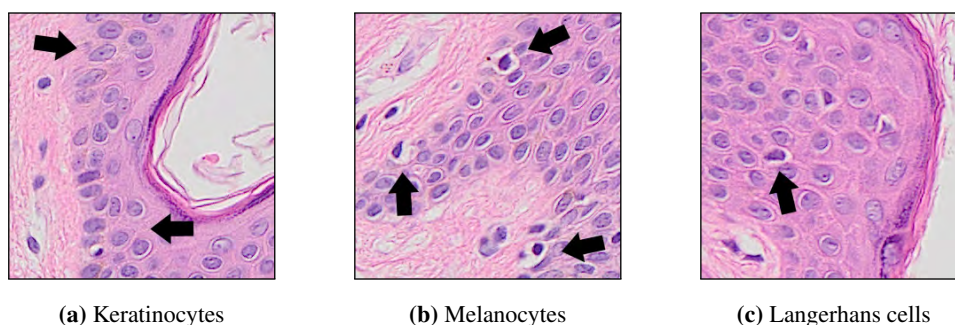


Figure 2.15. Selected cellular components

The downward projections of the epidermis between the underlying connective tissue are called rete ridges or retes (Fig. 2.16).

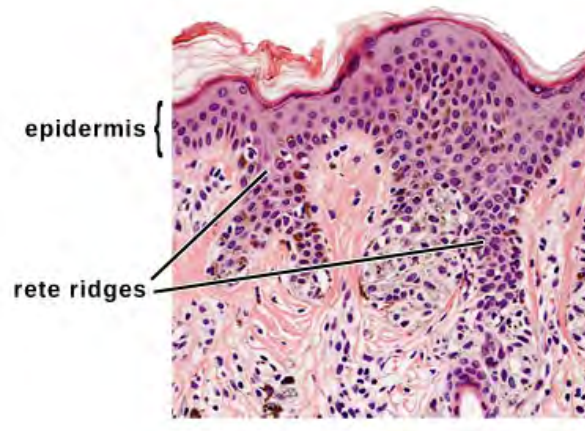
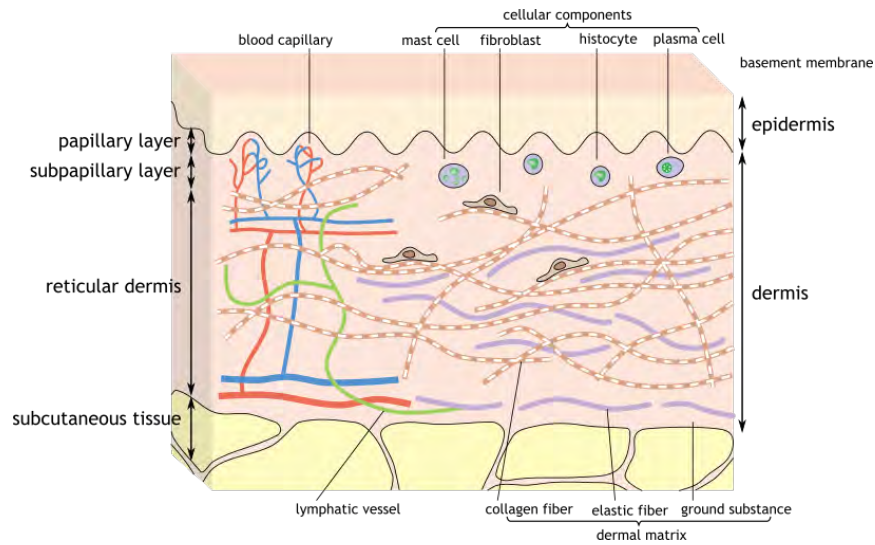


Figure 2.16. Rete ridges

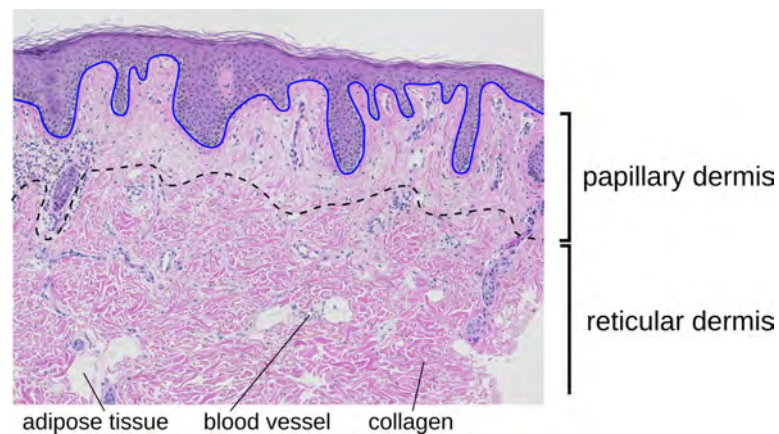
2.4.2. Dermis

The dermis is the structure beneath the epidermis, and the two are separated by the basal membrane. The dermis is approximately 15 to 40 times thicker than the epidermis. This region is irregularly arranged and filled mostly with connective tissue. It consists of three layers with no clear demarcation between them (Fig. 2.17a):

- **Stratum papillare** (papillary layer): The dermal area that projects into the intervals between the epidermal ridges; those raised irregular projections that interlace with the epidermal ridges of the epidermis are called dermal papillae. It consists mainly of a finely woven meshwork of collagen fibers.
- **Subpapillary layer**: The thin area underlying the epidermis, containing the same components as the papillary layer.
- **Stratum reticulare** (reticular layer): Accounts for the largest part of the dermis and has dense connective tissue comprising fiber components. The lower part comes into contact with the subcutaneous fatty tissue. There are blood vessels and nerves in some parts. The components of the dermis comprise the fibrous tissue and the dermal matrix formed by cells in the interstitial components. The major components mainly consist of collagen fibers (mainly types I and III), with smaller amounts of elastic fibers, and matrix. The integrity of the fibrous network varies with age and even within particular regions of the body. The matrix generally comprises the extra-cellular matrix and ground substance made up of proteoglycans and gelatin. Fibroblasts, macrophages, mast cells, plasma cells, vascular channels and nerves are common cellular components. The reticular dermis also contains other structures, such as appendages: sebaceous glands, sweat glands, and hair follicles.



(a) Schematic

(b) H&E ($\times 10$)**Figure 2.17.** Structure of the dermis.

The major components of each dermal layer are collagen fibers, elastic fibers, and matrix:

- **Collagen fibers.** Collagen fibers, formed from aggregations of thin fibrils, are the primary components of dermis, account for roughly 70–75% of the weight of dry dermis. The structure of collagen in dermis changes from thin and sparsely distributed fibers in the papillary and subpapillary layers to thick and densely distributed fibers in reticular layer (Fig. 2.17b). In adults, approximately 80–85% of dermal collagen fibers is type I collagen, 15% (mainly reticular fibers, which distribute in the perivascular regions) is type III collagen, and most of the remainder is thought to be type V collagen. The type III collagen fibers do not form thick fiber bundles.
- **Elastic fibers.** These fibres are highly elastic and impart elasticity to the skin. However, they cannot be differentiated from collagen fibers by H&E staining.
- **Ground substance (matrix).** Ground substance is an amorphous mixture of sugar and proteins, observed in between fibers and between cells in the dermis.

Selected common cellular components found especially in the reticular layer:

- **Fibroblasts:** Fibroblasts are thin spindle-shaped cells sparse in collagen fibers (Fig. 2.18a).
- **Histiocytes:** Histiocytes are broadly distributed in the connective tissue and mix together with fibroblasts (Fig. 2.18b).
- **Plasma cells:** Plasma cells are white blood cells that secrete large volumes of antibodies (Fig. 2.18c).

Apart from permanent cells there are also migratory cells, such as lymphocytes and other leukocytes.

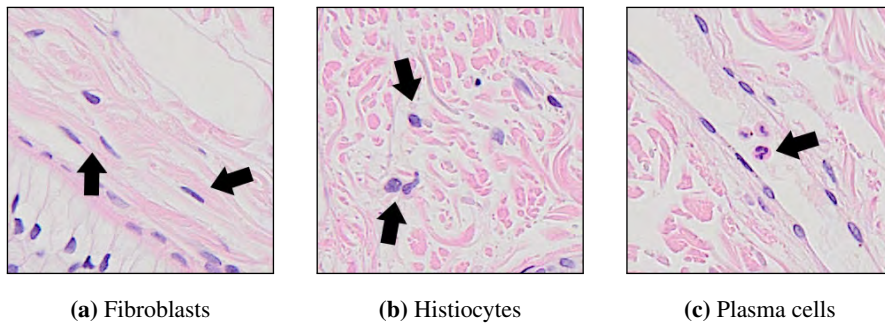


Figure 2.18. Selected cellular components

Dermis has a dense network of vascular channels and nerves:

- **Blood vessels:** Numerous branches of blood vessels distributed in skin are connected with each other forming two horizontal networks, the subcutaneous plexus and the subpapillary plexus, with arterioles ascending through the papillary layer forming capillary loops in the dermal papillaries before moving to venules (Figs. 2.17a and 2.19a). The peripheral regions of the eccrine glands and hair follicles in the growth stage are particularly richly supplied with vascular networks.
- **Lymphatic vessels:** Lymph vessels are distributed around the subpapillary layer and extend to the dermal and subcutaneous lymph vessels. They are partly ruptured, surrounded by loose collagen fibers and elastic fibers (Fig. 2.19b), and their structure is less regular than that of the blood vessels. The endothelial cells of lymph capillaries are thin and lack pericytes or lamina densa.
- **Nerve fibers:** In the dermal lower layers nerve fiber bundles are myelinated, i.e., covered with a membrane, but as they ascend through superficial dermis and peripheral appendages, they change to non-myelinated bundles and branch into many fibers (Figs. 2.19c and 2.19d).

2.4.3. Hypodermis

The hypodermis (subcutaneous tissue) is the layer between the dermis and the fascia, largely composed of fat cells. This region is rich in fiber bundles produced in the dermis and firmly connected with the fascia and periosteum through the subcutaneous tissue. The main component of the fat droplet is triglyceride. On H&E-stained images it appears transparent, as the content of fat cells is leached out during fixing and processing stages of the slide preparation process (and thus it absorbs neither hematoxylin nor eosin stain).

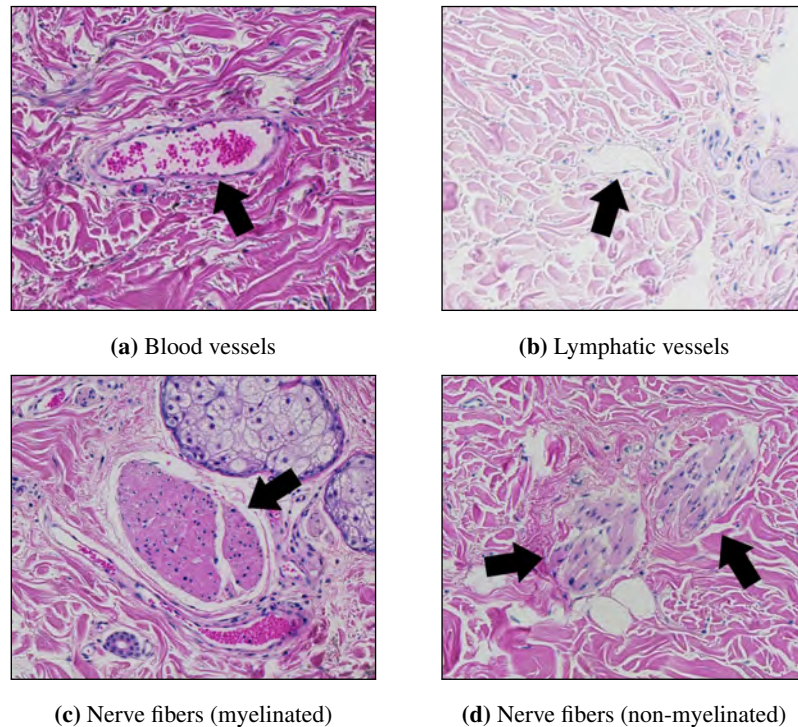


Figure 2.19. Vascular channels and nerves

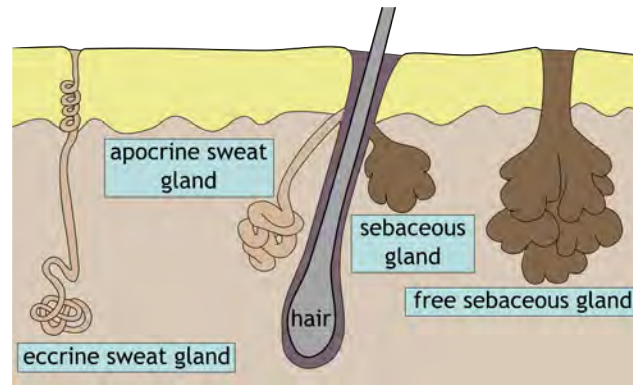
2.4.4. Appendages

Skin appendages (or adnexa) are skin-associated structures that serve a particular function including heat loss, lubrication, and contractility. Types of appendages most important for diagnosis of skin lesions include hair and glands (Fig. 2.20a).

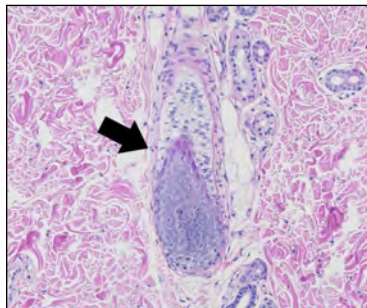
Hair apparatus. The hair apparatus is distributed throughout the skin except on the lips of the mouth. It consists of a hair and a hair follicle, the layer of tissue that encloses a hair and opens in a funnel shape. Hair follicles are aligned obliquely to the skin surface. Part of the hair follicle is slightly enlarged to form a hair bulge, above which sebaceous glands and apocrine glands are seen (Figs. 2.20c, 2.20b). During the growth stage the bottom of the hair root bulges out spherically forming a hair bulb.

Sebaceous glands. The sebaceous glands (Fig. 2.20d) are widely distributed throughout the skin, except in palms and soles and some mucous membranes.

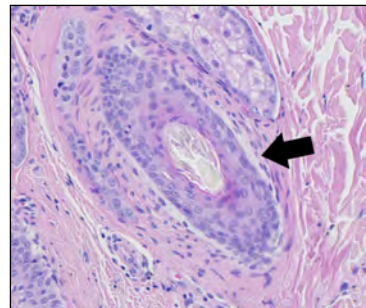
Sweat glands. Human sweat glands are either eccrine, distributed throughout most of the body, or apocrine, found at specific sites of the body. Both types are hair follicle-associated and consist of a secretory part and a sweat duct. The secretory parts are coiled and surrounded by fat tissues in the deep dermal layer and subcutaneous tissue. Eccrine sweat glands has two layers of secretory cells (Fig. 2.20e). The sweat duct ascends perpendicularly in the dermis (straight duct) through the coiled duct, which extends from the secretory area.



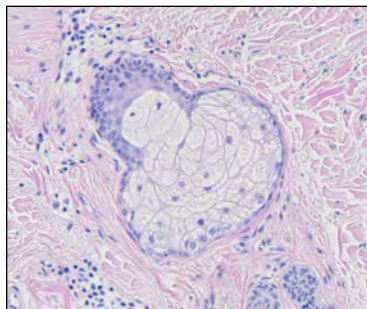
(a) Sweat glands, hair follicles and sebaceous glands.



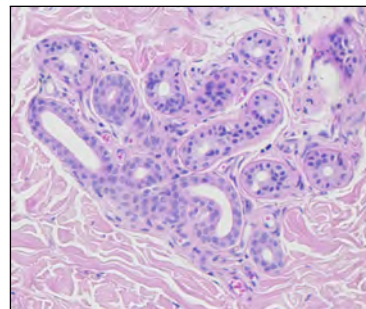
(b) Hair apparatus (longitudinal section)



(c) Hair apparatus (cross section)



(d) Sebaceous glands



(e) Sweat glands

Figure 2.20. Skin appendages

2.4.5. Morphometry of cells and tissue structures

Our methods rely on the morphometry of cells and tissue structures which can be found in the human skin. This section discusses and summarizes dimensions of the respective structures used in our study.

Epidermis. The diameter of keratinocytes varies from 8–25 μm whereas the diameter of an average keratinocyte is 12–16 μm [74]. The diameter of keratinocyte nucleus is $8.6 \pm 2.8 \mu\text{m}$ [75]. The mean thickness of cellular epidermis depends on the body site. For the typical body locations, the thinnest epidermis can be found in forearm dorsal, with its epidermal plate thickness varying between 45–65 μm [76, 77]. Therefore, we adopted the value of 40 μm as the minimum thickness of cellular epidermis. The mean epidermis-dermis thickness in typical body sites varies from 1.55 mm for tight to 2.54 mm for suprascapular area [78], although for thick skin it may reach up to 4 mm. The upper range of the 95% CI interval for the mean epidermis-dermis thickness in suprascapular area (2.58 mm) was adopted as the maximum

thickness of epidermis-dermis [78]. The typical thickness of stratum corneum is 10–30 μm , depending on the body site [79]. However, in pathologically changed epidermis the mean thickness of this layer may be up to 150 μm [80]. In our study we adopted the value of 100 μm as the maximum thickness of stratum corneum.

Based on those information, we have defined quantities used in image processing routines, which are summarized in Table 2.2.

Table 2.2. Values of morphometric parameters of tissue structures

Symbol	Description	Value
r_{KerMax}	maximum radius of a keratinocyte	12 μm
r_{KerAvg}	radius of an average keratinocyte	8 μm
r_{KerNuc}	radius of a keratinocyte nucleus	4 μm
A_{KerMax}	maximum area of a keratinocyte	πr_{KerMax}^2
A_{KerNuc}	area of a keratinocyte nucleus	πr_{KerNuc}^2
d_{EpMin}	minimum thickness of cellular epidermis	40 μm
$d_{\text{EpDermMax}}$	maximum thickness of epidermis-dermis	2.58 mm
p_{EpMin}	ratio of epidermis area to epidermis-dermis area	$d_{\text{EpMin}}/d_{\text{EpDermMax}}$
d_{SCAvg}	typical thickness of stratum corneum	30 μm
d_{SCMax}	maximum thickness of stratum corneum	100 μm

Dermis. The diameter of collagen fibers is quite variable, ranging from 2–15 μm [71, 81]. In abdominal skin the spacing between collagen bundles is 2–7 μm , whereas the percent fractional-area occupied by the collagen is 48–82% [81]. Elastic fibers are 1–3 μm in diameter [70]. Resting lymphocytes measure on average 8–10 μm in diameter and have a small, round nucleus that appears deeply basophilic because of the presence of numerous chromatin particles; they have only a narrow rim of cytoplasm that is difficult to delineate by conventional light microscopy. Since it is typically not possible by light microscopy to distinguish lymphocytes from monocytes in H&E-stained histologic sections, cells with a histologic appearance of lymphocytes are usually referred to as lymphoid cells (similarly, infiltrates likely to contain significant admixtures of both lymphocytes and monocytes are sometimes described as “lymphohistiocytic infiltrates”) [71].

2.5. Skin pathology

When observing a pathological specimen, it is necessary to identify the abnormality in the specimen by comparison with normal findings (Fig. 2.21a). Some fundamental terms for skin pathological changes and diseases are as follows [70]:

- **Acanthosis** (epidermal hyperplasia) describes thickening of the epidermis (Fig. 2.21b).

- **Epidermal atrophy** (epidermal hypoplasia) is caused by reduction of keratinocytes, which leads to thinning of the epidermis (Fig. 2.21c). Consequently, the papillary processes are diminished or even completely lost.
- **Hyperkeratosis** is abnormal thickening of stratum corneum (Fig. 2.21d).
- **Hypergranulosis** describes abnormal increase in the number of granular cell layers, to four or more layers (Fig. 2.21e).
- **Inflammatory cell infiltration** occurs when inflammatory cells infiltrate around the blood vessels (Fig. 2.21f). The principal infiltrating cells in melanocytic lesions are neutrophils, eosinophils, and lymphocytes.

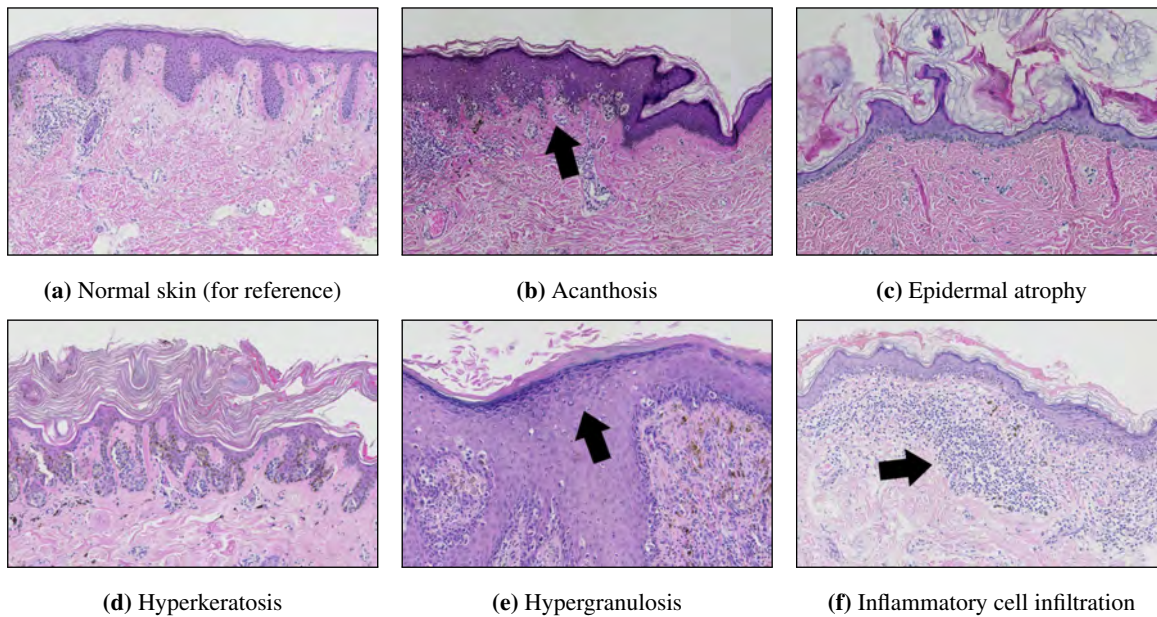


Figure 2.21. The selected fundamental abnormalities in skin specimens.

2.6. Skin melanocytic lesions

Skin melanocytic lesions are neoplasms derived from epidermal melanocytes. The two principal classes of skin melanocytic lesions are “benign” nevi (with no metastatic potential) and “malignant” melanoma (with a metastatic capacity proportional to its thickness).

In the context of melanocytes, the term “nevus” denotes the localized aggregation of nevus cells, i.e., cells derived from melanocytes, as benign neuroectodermal proliferations and/or neoplasms [82]. Nevus cells, arising as a result of proliferation of melanocytes at the dermal–epidermal junction, are typically grouped in nests [83]. Although the histogenesis of nevus cells remains the subject of debate, they are widely accepted as morphological variants of native melanocytes [84, 85, 86]. Nevi may be congenital or developed during lifetime, with the majority of them appearing during the first two decades of a person’s life [87]. Acquired melanocytic nevi are extremely common and most of them are benign with little malignant potential; only a very small proportion progress to melanoma [83]. The possible

pattern of evolution of pigmented nevi has been deduced by pathological examination of a large number of lesions removed at different ages [88, 89]. Moreover, it has been observed across many datasets that total numbers of nevi (especially dysplastic ones) constitute a major, if not the major, phenotypic risk indicator for all types of melanoma [90, 91].

Melanoma is the generic term for all malignant neoplasms derived from melanocytes, with most melanomas develop through an out-of-control progressive proliferation of melanocytes within the squamous epithelium [17]. Melanomas are the most aggressive and dangerous neoplasms, they grow fast and quickly metastasize to nearby lymph nodes and other organs.

In recent years, the sharp dichotomy between “benign” nevi and “malignant” melanoma started to be considered an oversimplification [16]. Although the great majority of melanocytic lesions can be confidently labeled as either nevi or melanoma, there exists a “gray zone,” which is composed of lesions that cannot be precisely classified into one of these groups. These distinct clinicopathological entities seem to correspond to low-grade forms of melanoma, in which distant metastasis is a rare but finite risk [92]. No histopathological criterion is entirely specific for nevi (all can be occasionally found in a melanoma) and some criteria are infrequent, or have low specificity, or suffer of high inter-observer variation. Moreover, some challenging lesions contain conflicting criteria suggesting opposite diagnoses. Therefore, criteria must be used in clusters: each single criterion must be evaluated and added to the others; the diagnosis of a nevus should be the outcome of the combination of clues to benignity found in the neoplasm. Apart from the “positive” criteria, dermatopathologists must also take into consideration the “negative” ones, i.e., the criteria that would suggest malignancy but are lacking. Additionally, criteria ought to be used for specific differential diagnosis, i.e., a specific form of nevus should be differentiated from a type of melanoma morphologically mimicking it: Spitz nevus vs. spitzoid melanoma, desmoplastic nevus vs. desmoplastic melanoma, etc. [16].

2.6.1. Benign lesions: Lentigo maligna

Lentigo maligna is a precursor to lentigo maligna melanoma, a potentially serious form of skin cancer. It occurs on the most sun-exposed surfaces, in sun damaged skin, so is generally found on the face or neck. It typically progresses slowly, over 5 to 20 years or longer, and can remain in a non-invasive form for years. Lentigo maligna is a proliferation of malignant melanocytes along the basal layer of the epidermis and within the hair follicle. One of the major histological criteria for the diagnosis of lentigo maligna, related to epidermal morphometry, is atrophy and effacement of the epidermis [17]. Examples of the lentigo are shown in Figure 2.22.

2.6.2. Dysplastic lesions: Junctional dysplastic nevus

The term “dysplastic nevus” was coined in 1978 by Wallace Clark and his colleagues [93], who proposed a hypothesis that a set of findings in nevi of members of families with a tendency to develop

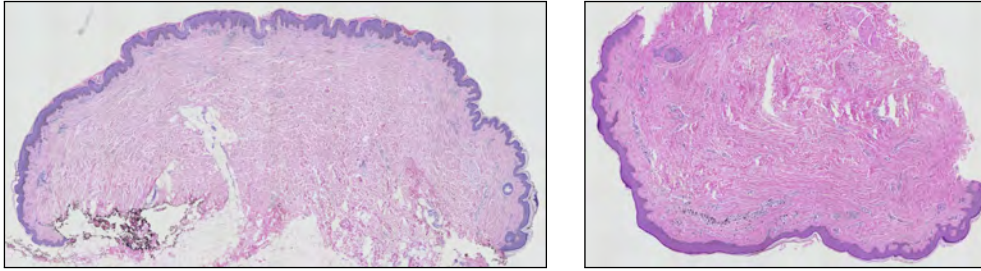


Figure 2.22. Examples of the lentigo

melanoma comprise a clinicopathological entity, representing an intermediate step between the “common nevus” and melanoma (hence a dysplastic nevus is often called a Clark nevus).

A junctional nevus is an intermediate stage of development between a banal common nevus and a dermal nevus. It comes into being when a single abnormal melanocyte begins to proliferate in a lentiginous pattern (i.e., forming a row) at the dermal epidermal junction (when the melanocytes are found both at the dermal epidermal junction and within the dermis, such a nevus is called a compound nevus).

In general, a junctional dysplastic nevus (also atypical nevus, nevus with architectural disorder and melanocytic atypia) consists of a lentiginous proliferation of a variable combination of single melanocytes and nevus cells in nests along the dermal-epidermal junction, although dysplasia may also extend into the epidermis above a dermal component. If present, nests are often irregular in size and shape and may “bridge” or join together (in common nevi the nests are roundish or regularly elongated, and typically positioned at the tips of rete ridges). The rete ridges in lentiginous areas are elongated, with bridging between adjacent rete ridges in some areas [92]. Some cells may contain finely granular melanin pigment. In 1985, on the conference sponsored by the World Health Organization (WHO), the following two major histological criteria for the diagnosis of dysplastic nevus were proposed [94]: (1) basilar proliferation of atypical melanocytes extending at least three rete ridges beyond the dermal nevocellular nevus; and (2) lentiginous or epithelioid cell pattern of the atypical intraepidermal melanocytic proliferation. Minor criteria include, among others, fusion of rete ridges. Histological dysplasia may be mild, moderate or severe. Examples of the junctional dysplastic nevus are shown in Figure 2.23.

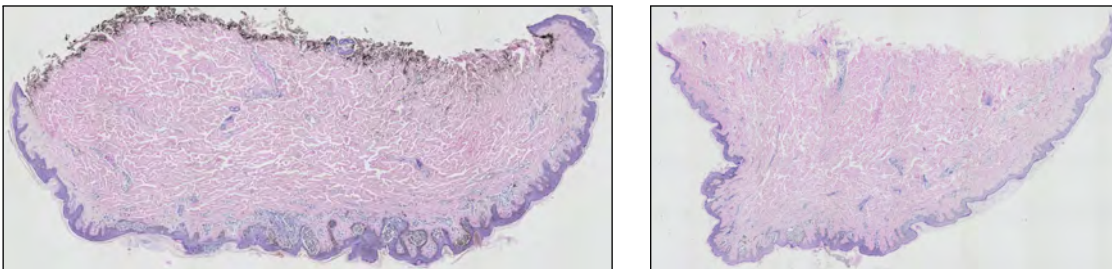


Figure 2.23. Examples of the junctional dysplastic nevus

There is no gold standard for the diagnosis of a dysplastic nevus [95]: at the “low end” of the spectrum are common nevi with some of the features seen in prototypical lesions, and at the “high end” are thin melanomas that histopathologically simulate such nevi. Nonetheless, the notion of dysplastic nevus as an

intermediate step in the tumor progression (between common nevus and melanoma) seem logical from the point of view of tumor biology [16].

2.6.3. Malignant lesions: Melanomas

Although the two most commonly diagnosed skin cancers are basal cell carcinoma and squamous cell carcinoma, which develop from the non-pigmented cells of the skin, the most aggressive and dangerous is melanoma. Melanoma originates in pigment producing cells called melanocytes and is less common but far more deadly than cancers mentioned above [96]. Melanomas are fast-growing and highly malignant tumors often spreading to nearby lymph nodes, lungs and brain (Fig. 2.24). Prognosis depends on the kind of melanoma, its size, location and depth of invasion. The most important parameter which predicts the stage of melanoma is the thickness of the examined lesion. Skin moles with the thickness less than 1 mm are nearly 100% curable [12]. Therefore, the aim of each clinician is to detect malignant melanomas when they are still small and thin.

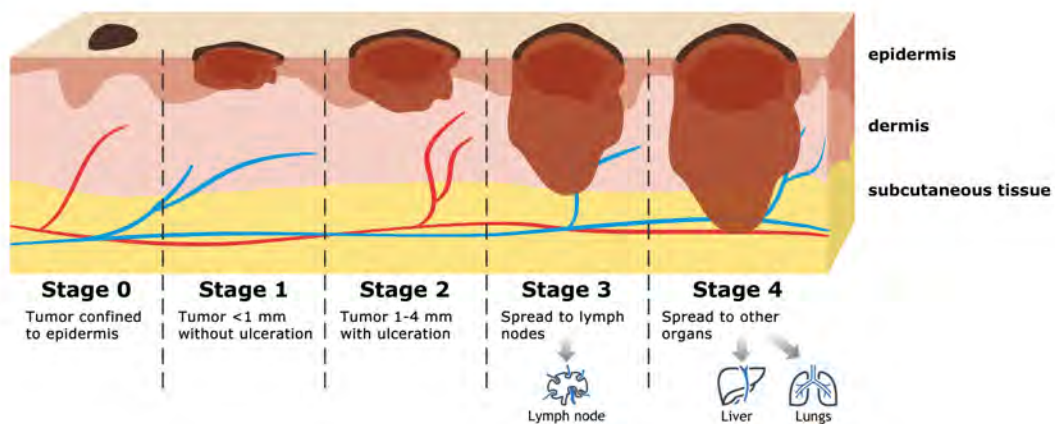


Figure 2.24. Presentation of five stages in malignant melanoma evolution process

Melanoma in situ is a form of melanoma entirely restricted to the epidermis, the dermoepidermal junction, and epithelial appendages. Melanoma in situ shows quite heterogeneous features and the three main diagnostic details typically observed in such lesions are: areas in which single melanocytes predominate, irregular distribution of junctional nests, and pagetoid spread of melanocytes (occasionally only one or two of these details are present in a lesion) [16]. Histopathological diagnosis of melanoma is difficult due to the lack of a single criterion that is specific for melanoma. The fundamental problem is especially true for melanoma in situ, as the currently used diagnostic criteria have been formulated based on the assessment of the intraepidermal component of invasive melanomas that have metastasized and therefore are not entirely representative (since tumor progression leads to changes in the characteristics of intraepithelial melanoma, these criteria may be different from that of “true” in situ lesions) [97]. Consequently, in many instances the diagnosis may be subjective because of differences in training and philosophy, and the experience of the observers: accordingly, the same lesion may be classified as dysplastic nevus with severe atypia by one pathologist and as melanoma in situ by another.

Superficial spreading melanoma is a term used to describe melanoma in a superficial, horizontal growth pattern, either in situ or initially invasive. The superficial spreading pattern of melanoma grows mostly horizontally in the epidermis, resulting in an asymmetric but vaguely elongated rectangular silhouette of SSM. Examples of superficially spreading melanomas are shown in Figure 2.25.

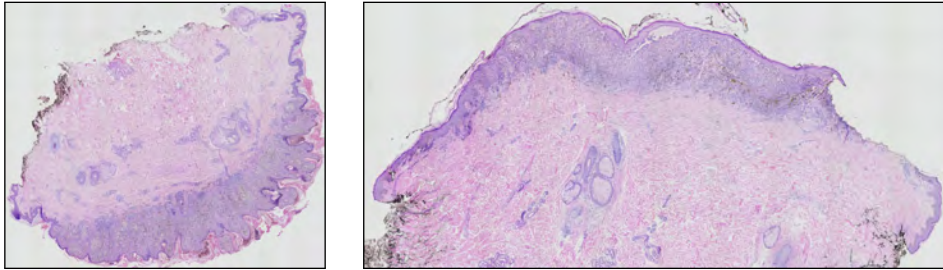


Figure 2.25. Examples of the superficially spreading melanoma (SSM)

Criteria for superficial spreading melanoma (SSM), which is the most common type of melanoma, and melanoma in the in situ stage include [16, 17]:

- **Epidermal plate:** The epidermis is irregularly altered, with hyper- and hypoplastic sections distributed in a disorganized fashion. The uniform atrophy of the epidermis and the presence of an elongated, dermoepidermal artifactual cleft are both indicators of melanoma.
- **Rete ridges:** Retses are usually widened or changed in another way. Nonetheless, occasionally, they may become completely effaced. Only exceptionally retes are thin and evenly elongated as in nevi.
- **Nests:** Large, irregularly shaped, confluent nests are unevenly distributed along the dermoepidermal junction, separated one from another by “skip areas” which are either free of melanocytes or with a lesser number of melanocytes arranged in a lentiginous pattern.
- **Infiltration:** There are variably dense infiltrates of lymphocytes and melanophages in the papillary dermis.

A key finding is that these changes are usually distributed unevenly.

Although some of the above-mentioned criteria are not directly related to epidermal morphometry, they should be considered when designing automatic image processing methods (e.g., the presence of abundant lymphocytic infiltration in dermis may affect the output of some stain deconvolution methods described in Section 2.3.4).

2.6.4. Differential diagnosis between benign and malignant lesions

The histological criteria currently used in the diagnosis of melanoma consist of the analysis of numerous features, such as: lesion’s asymmetry, morphometric features of epidermis, proliferation patterns of single melanocytes, cytological atypia, mitoses, and necrosis [16, 17].

According to [98], cytological atypia and asymmetry showed a high sensitivity (>90%), whereas suprabasal melanocytes showed a high specificity (>90%). Nonetheless, since many of the architectural characteristics appear to be more reproducible than assessment of cytological atypia, they take precedence over cytology as histological criteria [17].

Criteria related to the epidermal component are: uniformity and symmetry of an epidermal hyperplasia along the lesion, and elongation and thickening of rete ridges (Figs. 2.26 and 2.27). Clinical experience suggest the following relations between those criteria and selected classes of melanocytic lesions: (1) lesion symmetrical with respect to epidermal thickness (i.e., thickness is measured along the main axis of the epidermal region, on cross-sections orthogonal to that axis) usually indicates a nevus, whereas its asymmetry is often a sign of malignancy, (2) irregular hyperplasia is sometimes sighted in melanomas, (3) more uniform rete ridges are characteristic feature of nevi, (4) irregular alteration between elongation and effacement of rete ridges is often seen in malignant lesions, (5) although rete ridges of dysplastic nevi are often distorted, high percentage of the ordinary benign melanocytic naevi also showed some bridging of rete ridges [17]. The above-mentioned relations are considered to be particularly effective in differentiating between melanoma in situ (which shows quite heterogeneous features) and benign lesions (especially those resembling melanoma, like junctional nevus or lentigo simplex) [16].

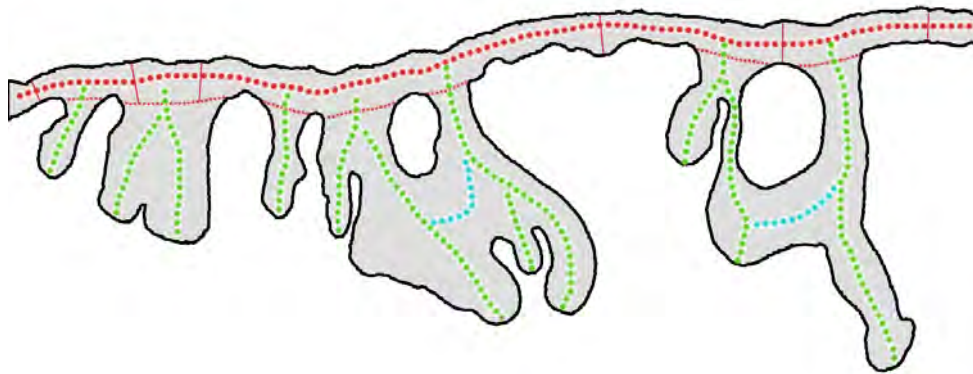
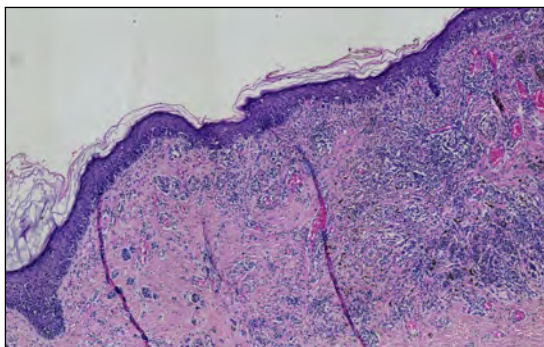
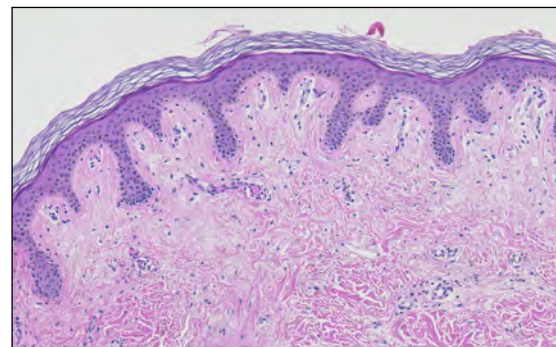


Figure 2.26. A diagram of the structural features of the epidermis: main axis of the epidermal region (large red dots), boundary of the epidermal “base” (small red dots) used to measure thickness of the epidermis (red bars, perpendicular to the main axis), main axes of rete ridges (green), bridging of adjacent rete ridges (blue). This epidermal section is asymmetric – its right half has considerably more elongated rete ridges than its left half.



(a) Melanoma in situ (malignant)



(b) Junctional dysplastic nevus (benign)

Figure 2.27. A comparison of the structure of an epidermal component: (a) melanoma in situ, effacement of rete ridges; (b) junctional dysplastic nevus, uniformly elongated rete ridges.

The problem is that those criteria are vaguely defined and present low specificity and/or sensitivity [16]. There are no comprehensive studies on the efficacy of particular criteria [17]; the existing study

covers only a narrow subset of all routinely used criteria (which does not include morphometric features of epidermal component) [98].

2.7. Melanoma misdiagnosis problem

The histopathological examination constitutes the gold standard for classifying a melanocytic skin lesion as malignant. Although there exist other forms of examination (e.g., dermatoscopy), they do not yield as high diagnostic confidence in contrast to a histopathological examination (e.g., while expert dermatologists may arrive at 90% sensitivity and 59% specificity in skin lesion diagnosis, for young inexperienced dermatologists and family physicians these figures drop significantly till around 62–63% for general practitioners [99, 100]).

However, even in case of the histopathological examination there is often no concordance between pathologists regarding the diagnosis of a given lesion, and discordance in the diagnosis of melanoma poses a serious problem as it has important consequences to the patient depending on the nature of the lesion. Farmer et al. [101] analyzed the performance of a panel of 8 expert dermatopathologists in the interpretation of 37 cases of cutaneous melanocytic tumors exhibiting “classic features” as benign, malignant, or indeterminate. They found out that in 38% of cases two or more pathologists has discordant interpretations. Lodha et al. [2] examined the discordance among senior dermatopathologists in evaluating 172 cases of difficult melanocytic neoplasms (most cases were reviewed by two dermatopathologists). In 25.2% of cases the diagnostic discordance was extreme (“definite melanoma vs. definite nevus”) and in total in 36.4% of all cases the diagnoses were significantly discordant (i.e., one specialist claimed that a lesion is more likely benign while the other one that is more likely malignant). Troxel [102] studied 335 pathology malpractice claims and found out that 13% of them were false-negative melanoma claims (it had highest percentage in the category of systematic errors). Shoo et al. [14] compared the initial diagnosis rendered to 392 patients by an outside dermatopathologist or surgical pathologist with a subsequent diagnosis rendered after re-evaluation by routine histopathologic examination. The discordance rate of melanomas and nevi between the two was 14.3%.

The above-mentioned studies show that the misdiagnosis rate of melanoma may be as high as 25% and there are several main reasons behind this problem: (1) less common melanoma subtypes have different histopathological features than superficial spreading melanoma (the most common type of melanoma – 40-70% of all diagnosed cases, depending on the region of the world) and thus may be more difficult to diagnose, (2) lesions are evaluated not by dermatopathologist or surgical pathologist but by general histopathologists (who lack an in-depth knowledge of nuances between less common types of melanoma), (3) some uncommon form of melanoma (the so-called nevoid melanoma) resembles a common benign melanocytic nevus on histological examination, and (4) the diagnosis is mainly based on experience and intuition of a given pathologist (therefore it is highly subjective) [3, 14, 103, 104].

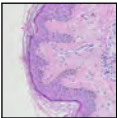

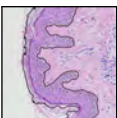

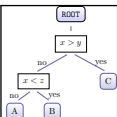
A false-negative diagnosis of melanoma (i.e., misdiagnosing a malignant lesion as a benign one) is the most serious issue as it delays the beginning of treatment what leads to complications and even patient’s

death. Figures show that a false-negative diagnosis of melanoma is a continuing problem for pathologist – in the United States it has become the single most common reason for filing a malpractice claim against a pathologist (70–95% of all cases) [102, 105].

These obstacles encourage researchers to try to implement and build computer-aided diagnosis (CAD) systems for automated diagnosis of melanoma aimed at increasing the specificity and sensitivity of the assessment of skin melanocytic lesions and at its simplification.

3. Preface to the proposed algorithms

The proposed automatic diagnostic algorithm consists of the following stages:

(Input)	A high-resolution RGB whole slide image of a H&E-stained skin specimen	
↓		
Tissue segmentation	Segmentation is done indirectly, by detecting the slide background regions, using statistical analysis of color distribution, color thresholding and reconstruction in CIELAB color space (described in Chapter 4).	
↓		
Epidermis segmentation	Segmentation is based on information about shape and distribution of slide background regions in an image and information about distribution and concentration of histological stains (described in Chapter 5).	
↓		
Retes segmentation	Segmentation of rete ridges in the previously segmented epidermis and measurement of their basic morphometric features is done mainly by analyzing the curvature of the epidermal boundary and the skeleton of the epidermal region in the graph representation (described in Chapter 6).	
↓		
Feature extraction & classification	A machine learning classification algorithm makes a diagnosis based on features describing the overall epidermal morphometry, computed from information extracted in the previous stage (described in Chapter 7).	
↓		
(Output)	diagnosis: lentigo / dysplastic nevus / melanoma	

In the subsequent sections we summarized the state-of-the-art literature and challenges related to each stage of the proposed diagnostic algorithm, briefly described the outline of methods proposed by the Author for each stage, and characterized datasets used in our study.

3.1. Related works in the field

This section gives a brief introduction to the subject matter of each of the system's modules, discusses key challenges to be considered when designing its algorithms, summarizes the most important state-of-the-art methods, and provides an outline of the method proposed by the Author.

3.1.1. Tissue segmentation

The accurate tissue segmentation is usually an obligatory first step in computerized analysis of WSIs, e.g., to improve the accuracy of stain deconvolution methods or to perform morphological analysis of the tissue structure [24, 58]. It is a challenging task as both inter- and intra-image color variations must be taken into account when analyzing histopathological images, as mentioned in Section 2.3.4.

In recent years a few general-purpose approaches to tissue segmentation in low-resolution WSIs were proposed. Bug et al. proposed the FESI method based on global thresholding at the mean value of the Gaussian blurred Laplacian of the greyscale image, refined by flood filling from identified background points [106]. Bándi et al. used a deep learning approach (a fully convolutional neural network and a U-net), which managed to detect large void spaces in the tissue but failed to segment medium- and small-sized pores (as mentioned by the authors in [107]). Additionally, a general-purpose GrabCut (GC) method is often used for image regions segmentation [108]. These tissue segmentation methods are intended either to be used in whole-slide scanners, where a coarse foreground estimation (at low resolution of approximately 4–8 $\mu\text{m}/\text{px}$) is enough to determine regions to be scanned and the proper focus depth for each of them, or to segment only large void regions within tissue. Nonetheless, those methods fail when applied to a skin specimen, since it contains numerous little void spaces between bundles of collagen fibers in dermis and frequently a big section of (colorless) adipose tissue as shown in Figure 4.1. The statistical comparison of the performance of all above-mentioned methods is presented in Section 4.2.3.

The approach proposed by the Author is based on color statistical analysis, CIELAB color thresholding, and binary morphology. As it is dedicated for H&E staining only, domain-specific knowledge about color properties observed in WSIs could be incorporated into the solution. Therefore it provides fine-grained, reliable, and easy to implement automatic foreground selection, tolerant to the slide magnification, illumination conditions, artifacts and noise.

3.1.2. Epidermis segmentation

The morphometric and cytologic features of the epidermal area are key factors considered when diagnosing skin conditions and during skin tissue grading, especially if the lesion involves inflammation [16,

109]. For example, symmetry and regularity of epidermal thickness are diagnostic clues useful to differentiate between (benign) nevi and (malignant) melanoma. Similarly, cellular atypia of keratinocytes typically indicates either squamous cell carcinoma or basal cell carcinoma, the two most common types of skin cancer.

Although the topic of automatic epidermis segmentation is not yet sufficiently explored, a few computerized techniques have already been proposed. Mokhtari et al. [110] suggested a method based on the contrast limited adaptive histogram equalization (CLAHE). Firstly, the noise in image is removed by applying morphological closing; then, the CLAHE is performed in order to adjust local contrast of the image; finally, the epidermis area is obtained using global thresholding. Haggerty et al. [111] proposed a technique based on both contrast enhancement and thresholding (henceforth referred to as the CET technique). It applies global thresholding on an image with its contrast enhanced (the enhancement involves combining a greyscale image and the b^* channel in CIELAB color space into the resulting image). Lu et al. [112] presented a technique based on global thresholding and shape analysis (henceforth referred to as the GTSA technique). Firstly, to obtain a coarse segmentation the global thresholding and shape analysis is used. Then, a template matching method with adaptive template intensity value is applied. Lastly, the probability density function of the response value image is used to calculate the threshold. Xu et al. [113] proposed an algorithm consisting of two steps. In the initial step a coarse epidermis segmentation on the WSI is obtained. Then, the thickness of this roughly delimited epidermis is measured along the main axis of the lesion region. Region exhibiting excessive thickness are analyzed once again to perform the segmentation process more precisely. Henceforth, this algorithm is further referred to as the THM technique. All the above-mentioned epidermis segmentation methods are based primarily on the combination of the global thresholding technique with area and shape analysis. Each of those methods makes substantial assumptions about color and contrast of WSI images. However, as mentioned in Section 2.2.3, inter- and intra-image variations in staining and illumination are quite common. Moreover, stain distribution will be strongly affected by the presence of certain pathological changes discussed in Section 2.5. Consequently, these assumptions are often not met when it comes to analyzing data from different sources and such segmentation methods tend to incorrectly include large false positive dermal regions in the resulting segmentation mask. The bulk of those false positive regions are dark skin components, including appendages, nests of melanocytes, and lymphocytic infiltration. In extreme cases, the aforementioned variations in image color may lead to a general failure of a given segmentation algorithm.

In the light of the above-mentioned limitations, a novel approach towards epidermis segmentation has been proposed, based on the method published by the Author in [24]. The initial segmentation is based on the analysis of shape and distribution of void regions in a tissue. It is then refined by analyzing the concentrations of hematoxylin and eosin stains in the tissue. In the final step, shape and location constraints are applied on the segmented region to obtain the actual epidermal region.

The proposed method strongly relies on domain-specific knowledge of cell and tissue morphology and their biochemical properties. The porosity analysis stage is based exclusively on the analysis of void spaces; in particular in its morphological operations it utilizes only such well-defined quantities as

radii of epidermal cells and their nuclei. The stain concentrations analysis stage is based on the analysis of normalized stain concentrations maps determined automatically for each individual image. It also utilizes knowledge about the way certain tissue structures absorb both hematoxylin and eosin. Therefore, this technique is robust to differences in staining protocols, stain absorption within individual images, and to illumination variability. Moreover, it does not require any parameter selection from its users, as the relevant values are either computed from the slide image or predefined based on domain-specific knowledge.

To sum up, the novelty of the proposed method is in its full automation, robustness to color variations in staining, and its solid basis in skin histology and dermatopathology. The proposed method can be used, as a preliminary step, by higher-level algorithms used for the analysis on human skin, e.g., for the detection of epidermal melanocytes or for the measurement of epidermal morphometry. Additionally, some of the proposed techniques (e.g., for measuring epidermis morphometry) may easily be refined and used in different fields of computer vision, such as general object analysis and recognition.

3.1.3. Epidermis morphometry

The epidermal area (Fig. 3.1) is an especially important target of examination when diagnosing skin conditions, especially melanocytic lesions. Its morphometric and cytologic features are crucial factors considered when grading a skin specimen [16, 109].

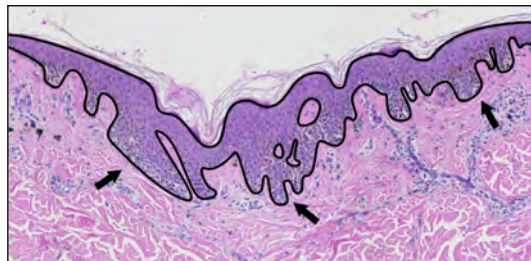
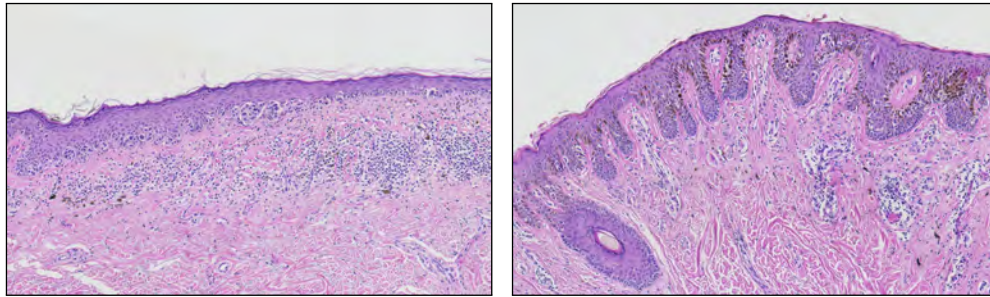


Figure 3.1. The epidermis (marked with the solid line) often has a complicated morphometry, with numerous rete ridges (some of them are marked with an arrow).

The main histopathological criteria related to the morphometry of epidermal component, which are currently used for melanoma diagnosis, include lesion's general (a)symmetry, uniformity and symmetry of epidermal hyperplasia along the lesion, and elongation and thickening of rete ridges [16, 17]. The epidermal thickness is measured along the sections of epidermis between bases of adjacent retes, perpendicularly to the main axis of the epidermal base (the epidermis main axis is the centerline of the epidermis base, i.e., of the epidermal region with retes passed over). A comparison of a typical epidermal morphometry in benign and malignant lesions is shown in Figure 3.2.

In this thesis a novel method for automatic determination of the location, base width, length, and height of retes in a segmented epidermis has been proposed. These information can help develop complex indexes describing the epidermal morphometry (as described in Chapter 7), which could be used by CAD systems in digital dermatopathology, or help in the cognitive analysis of lesions [114, 115]. **To the best**



(a) Melanoma in situ (malignant)

(b) Junctional dysplastic nevus (benign)

Figure 3.2. Comparison of the epidermal component structure: (a) melanoma *in situ* – effacement of rete ridges, irregular epidermal thickness; (b) junctional dysplastic nevus – uniformly elongated rete ridges, regular epidermal thickness.

of Author’s knowledge the proposed method is the first one to detect individual rete ridges and measure their morphometric parameters.

3.2. Datasets specification

To obtain the WSIs we established scientific cooperation with the University Hospital in Krakow and with the Chair of Pathomorphology of Jagiellonian University Medical College. Additionally, the Author was granted permission to use data available in public open image repositories of University of Michigan and University of British Columbia.

Throughout this study we used whole slide images of skin lesions acquired from three different laboratories and captured using four different devices (overview in Table 3.1):

- The JUMC-SS dataset was obtained from Jagiellonian University Medical College. The original images were captured under 10× magnification (0.44 μm/px) on Axio Scan.Z1 slide scanner and saved into uncompressed TIFF files.
- The JUMC-AM dataset was obtained from Jagiellonian University Medical College. The original images were captured under 20× magnification (0.345 μm/px) using Olympus BX51 light microscope equipped with Pike F505C VC50 digital camera and saved into uncompressed TIFF files.
- The UMch dataset was obtained from the University of Michigan Virtual Slide Box [116]. The original images were captured under 40× magnification (0.25 μm/px) on Aperio AT2 slide scanner and saved into JPEG2000 format at a quality level of 70.
- The UBC dataset was obtained from the Virtual Slidebox of University of British Columbia [117]. The original images were captured under 40× magnification (0.25 μm/px) on Aperio ScanScope CS² slide scanner and saved into JPEG format at a quality level of 30.

JUMC-SS and JUMC-AM datasets included cases of selected melanocytic lesions: lentigo maligna, junctional dysplastic nevus, melanoma in situ, and superficial spreading melanoma. UMch and UBC datasets included cases of chronic and acute conditions characterized by inflammatory skin, in particular:

Table 3.1. Parameters of WSI datasets used in this study. Origin of slides: JUMC-AM, JUMC-SS – Jagiellonian University Medical College; UMch – University of Michigan [116]; UBC – University of British Columbia [117].

Set ID	Mag.	$\mu\text{m}/\text{px}$	Image format	Device
JUMC-SS	10 \times	0.44	TIFF	Axio Scan.Z1
JUMC-AM	20 \times	0.345	TIFF	Olympus BX51 + Pike F505C VC50
UMch	40 \times	0.25	JPEG2000 q70	Aperio AT2
UBC	40 \times	0.25	JPEG q30	Aperio ScanScope CS ²

dermatitis, lichen, psoriasis, skin carcinoma, and skin nevus. For malignant lesions, we focused mainly on superficial spreading melanoma, as it constitutes 67% of all melanoma cases diagnosed nowadays [17].

The images differed in specimens' staining quality as well as original magnification, and were captured using either an automatic microscope or a whole-slide scanner, which allowed us to cover most of the variations which one would encounter in regular clinical practice. All the histological sections used in the evaluation were prepared from formalin-fixed paraffin-embedded tissue blocks of skin biopsies (each section was about 4 μm thick) stained with H&E using an automated stainer. The size of original WSIs varied from 3000 \times 1000 to 20000 \times 30000 pixels. Examples of WSI images from each of the datasets used in this study are shown in Figure 2.11.

To verify each stage of the proposed diagnostic system, WSIs were paired with the following diagnostic information (the ground truth) prepared manually by an experienced dermatopathologist: tissue segmentation (a binary mask), epidermis segmentation (a binary mask), information about each individual rete (the approximated location of its base endpoints and its length), and the diagnosis. Binary masks were prepared using GIMP image processing program and information about retes were provided using the "ROI Manager" functionality of ImageJ image processing program. The tissue segmentation stage was tested on 60 images from all four datasets (30 from JUMC-SS, 10 from JUMC-AM, 10 from UMch, and 10 from UBC dataset). The epidermis segmentation stage was verified using all images from JUMC-SS dataset. The retes segmentation stage was tested on 25 images from JUMC-SS dataset, which contained in total nearly 1000 manually segmented retes (manually annotating all 75 images would yield little advantages compared to the amount of effort needed to prepare the ground truth). The output of the complete diagnostic system (i.e., the predicted diagnosis for a WSI image) was evaluated on the whole JUMC-SS dataset.

4. Tissue segmentation

The accurate tissue segmentation is usually an obligatory initial step in computerized analysis of WSIs, e.g., to improve the accuracy of stain deconvolution methods or to perform morphological analysis of the tissue structure [24, 58]. It is a challenging task as both inter- and intra-image color variations must be taken into account when analyzing histopathological images, as mentioned in Section 2.3.4.

In recent years a few general-purpose approaches to tissue segmentation in low-resolution WSIs have been proposed. Bug et al. proposed the FESI method based on global thresholding at the mean value of the Gaussian blurred Laplacian of the grayscale image, refined by flood filling from identified background points [106]. Bándi et al. used a deep learning approach (a fully convolutional neural network and a U-Net), which managed to detect large void spaces in the tissue but failed to segment medium- and small-sized pores (as mentioned by the authors in [107]). Additionally, a general-purpose GrabCut (GC) method is often used for image regions segmentation [108]. These tissue segmentation methods are intended either to be used in whole-slide scanners, where a coarse foreground estimation (at low resolution of approximately 4–8 $\mu\text{m}/\text{px}$) is enough to determine regions to be scanned and the proper focus depth for each of them, or to segment only large void regions within tissue. Nonetheless, those methods fail when applied to a skin specimen, since it contains numerous little void spaces between bundles of collagen fibers in dermis and frequently a big section of (colorless) adipose tissue as shown in Figure 4.1.

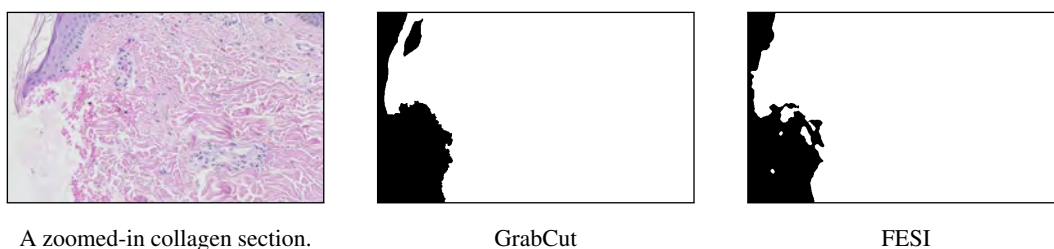


Figure 4.1. Performance of two existing tissue segmentation methods, GrabCut and FESI, on a collagen section in high resolution (0.25 $\mu\text{m}/\text{px}$): neither of these methods is capable of accurately segmenting the void space within collagen bundles.

The approach we propose is based on color statistical analysis, CIELAB color thresholding, and binary morphology. As we dedicated it for H&E staining only, we were able to incorporate domain-specific knowledge about color properties observed in WSIs. Therefore, it provides fine-grained, reliable, and easy to implement automatic foreground selection, tolerant to the slide magnification, illumination conditions, artifacts and noise.

4.1. Algorithm for tissue segmentation

Our method is based on the assumption that background pixels are similar to the most frequently occurring almost-achromatic color: both pixels close to that color as well as having similar chromaticity and higher intensity are considered slide background. The pipeline of our method is shown in Figure 4.2.

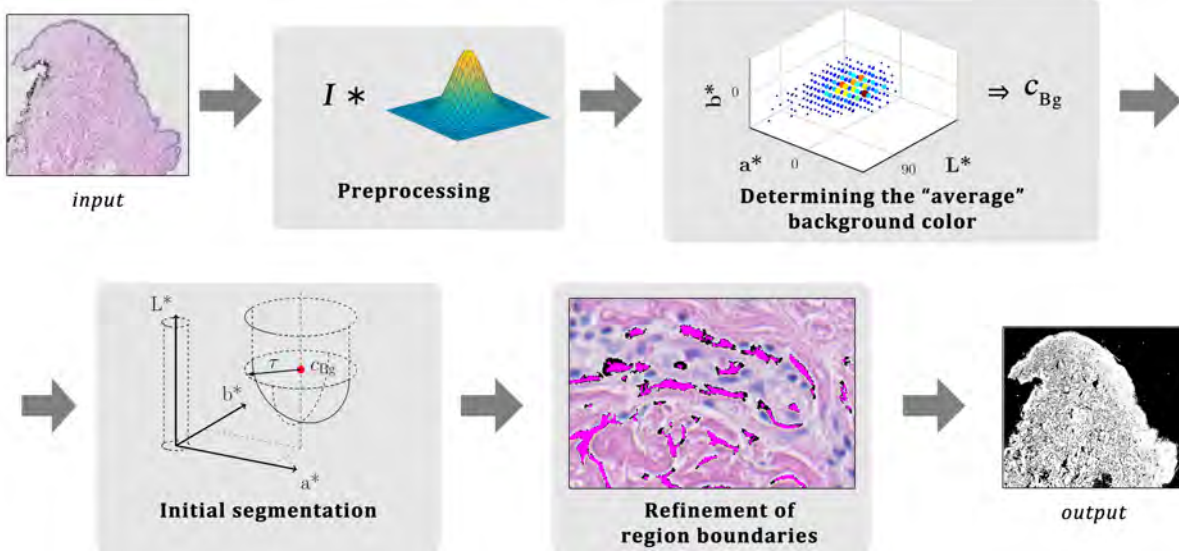


Figure 4.2. The pipeline of the proposed method

4.1.1. Preprocessing

Since the WSIs are taken using a CCD sensor, in the preprocessing step we reduce the effect of CCD noise by slightly smoothing the input RGB image using a Gaussian filter with $\sigma = 0.5$ and kernel size of 3 px (the values have been chosen experimentally).

4.1.2. CIELAB colorspace metrics

Our approach uses color thresholding in the CIELAB color space. In accordance with the design objective of a uniform color space, the color distance ΔE^* is intended to be perceptually uniform, i.e., if CIELAB is ideal, the same amount of numerical change in color component values corresponds to about the same amount of visually perceived change.

In our method we used the CIEDE2000 color-difference formula published by the International Commission on Illumination (CIE) in 2001 [118]. Given a pair of color values in CIELAB space $c_1 = (L_1^*, a_1^*, b_1^*)$ and $c_2 = (L_2^*, a_2^*, b_2^*)$, we denote the CIEDE2000 color difference between them as follows:

$$\Delta E_{00}^*(c_1, c_2) = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2} + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H} \quad (4.1)$$

where $\Delta L'$, $\Delta C'$, and $\Delta H'$ are the CIELAB metric lightness, chroma, and hue differences, respectively, calculated between the two color samples, R_T is an interaction term between chroma and hue differences, and S_L , S_C , and S_H are the weighting functions for the lightness, chroma, and hue components,

respectively. The k_L , k_C , and k_H values are factors to be adjusted according to such parameters as textures, backgrounds, separations, etc., for the lightness, chroma, and hue components, respectively (in our method we adopted $k_L = k_C = k_H = 1$). The exact formulation and implementation of all the aforementioned parameters can be found in [119].

Additionally, we defined the chroma-difference formula based on (4.1), which quantifies only chroma-related relationships:

$$\Delta E_{00CH}^*(c_1, c_2) = \sqrt{\left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}}. \quad (4.2)$$

Finally, the chroma (relative saturation) C^* of a color $c_P = (L_P^*, a_P^*, b_P^*)$ in the CIELAB color space is defined according to [120] as:

$$C_P^* = \sqrt{(a_P^*)^2 + (b_P^*)^2}. \quad (4.3)$$

4.1.3. Determining the “mean” background color

To determine the “mean” background color c_{Bg} , we firstly construct a 3D histogram of CIELAB values of nearly achromatic image pixels, i.e., pixels satisfying the condition $C_p^* < \tau_{Hist}$ (Fig. 4.3). The $JND \approx 2.3$ stands for a “just noticeable difference”, as proposed by Mahy et al. [121]. Then, from all non-zero bins $q\%$ largest bins is selected and the weighted mean of the CIELAB coordinates of their centers is computed (the share of each selected bin in the sum of sizes of all selected bins is taken as its weight). The threshold value $\tau_{Hist} = 4.075 \approx 1.77JND$ and the value $q = 0.5\%$ have been determined using the sequential quadratic programming (SQP) approach (see Section 4.2.2 for details). In order to maintain balance between method’s accuracy and robustness to noise, the histogram contains only pixels from the subspace of CIELAB color space for $L^* \in \langle 75, 100 \rangle$, $a^* \in \langle -10, 10 \rangle$ and $b^* \in \langle -10, 10 \rangle$, whereas the subspace is partitioned into $200 \times 256 \times 256$ bins.

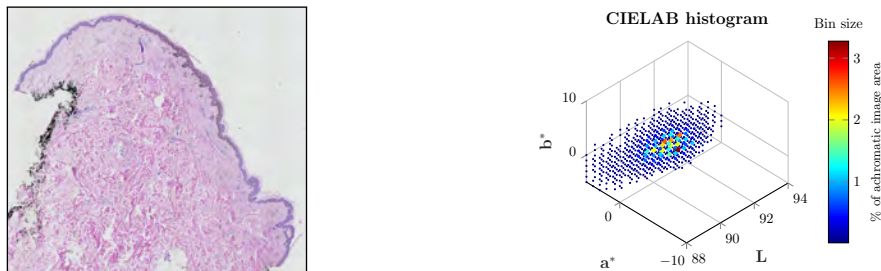


Figure 4.3. A WSI and its histogram of CIELAB values of nearly achromatic image pixels.

4.1.4. Initial background segmentation

After the “mean” background color is determined, we perform color thresholding in the CIELAB color space, based on CIEDE2000 color distance. Let us define the following conditions:

$$\begin{aligned} \text{Cond}_1(c_P, \tau) = & (\Delta E_{00}^*(c_P, c_{Bg}) < \tau) \\ & \vee (\Delta E_{00CH}^*(c_P, c_{Bg}) < \tau \wedge L_P^* > L_{Bg}^*) \end{aligned} \quad (4.4)$$

$$\text{Cond}_2(c_P) = L_P^* < \text{JND} \quad (4.5)$$

where c_P is the color of a pixel P and τ is a color distance threshold.

Pixels satisfying the condition Cond_1 either have color similar to the “mean” slide background color c_{Bg} or are brighter than c_{Bg} but have similar chromaticity. The condition Cond_2 segments those pixels which are nearly achromatic (disregarding their brightness). Such pixels usually appear in slide artifacts, like folded adhesive film or air bubbles under the coverslip. As H&E stains the whole specimen (i.e., all its tissue structures), we can without risk assume that grayish regions form part of the background. Figure 4.4 shows the CIELAB subspace restricted by conditions Cond_1 and Cond_2 .

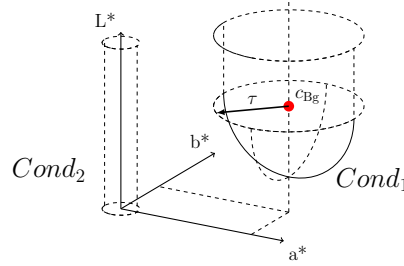


Figure 4.4. The CIELAB subspace restricted by conditions Cond_1 and Cond_2 .

We obtain the initial background mask M_{Bg}^0 by performing hysteresis thresholding (conditions are applied pixel-wise for each pixel P):

$$M_{\text{Core}} = \text{Cond}_1(c_P, 1.32\text{JND}) \cup \text{Cond}_2(c_P) \quad (4.6)$$

$$M_{\text{Base}} = \text{Cond}_1(c_P, 3.06\text{JND}) \cup \text{Cond}_2(c_P) \quad (4.7)$$

$$\begin{aligned} M_{\text{Constr}} = & (\rho_{\text{neighbors}}(P, M_{\text{Core}}, 10) > 2\%) \\ & \cup (M_{\text{Core}} \oplus \text{SE}_D(1)) \end{aligned} \quad (4.8)$$

$$M_{\text{Bg}}^0 = \left(M_{\text{Core}} \xrightarrow{8} M_{\text{Core}} \cup M_{\text{Base}} \cap M_{\text{Constr}} \right) \quad (4.9)$$

where $\rho_{\text{neighbors}}(P, M, r)$ is the share of ones in the mask M in P 's neighborhood of radius r pixels.

The threshold values $\tau_{\text{Core}} = 3.047 \approx 1.32\text{JND}$ and $\tau_{\text{Base}} = 7.049 \approx 3.06\text{JND}$ in Cond_1 in (4.6) and (4.7) have been determined using the sequential quadratic programming (SQP) approach (see Section 4.2.2 for details). The M_{Base} mask is further restricted with M_{Constr} to both limit the number of false-positives due to noise and to emphasize the fact that a pixel is more likely to constitute part of the background if it is neighboring a big “core” background region than if it is neighboring a tiny one. Figure 4.5 shows the visualization of pixels constituting the M_{Base} mask, whereas the M_{Core} mask overlaid on M_{Bg}^0 mask is shown in Figure 4.6.

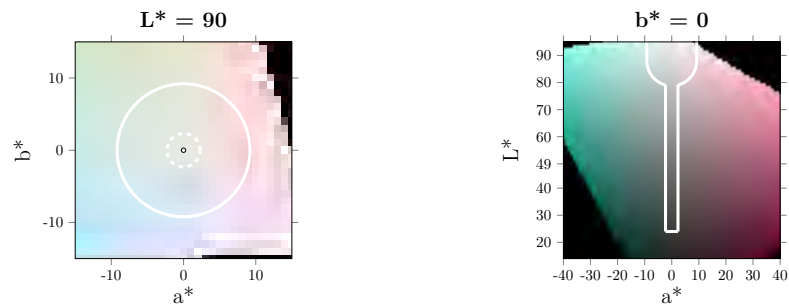


Figure 4.5. The region circumscribed by M_{Base} mask conditions on a^*b^* and L^*a^* color planes for $L^* = 90$ and $b^* = 0$, respectively (for $c_{\text{Bg}} = (88, 0, 0)$).

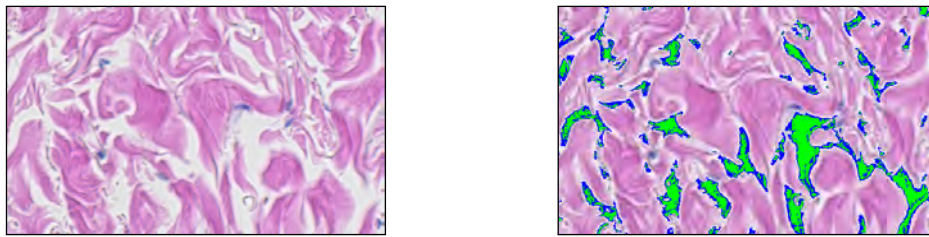


Figure 4.6. M_{Core} mask (green) overlaid on M_{Bg}^0 mask (blue).

To reduce the number of small segmentation artifacts, we subsequently both flood-fill those holes in M_{Bg}^0 mask which are smaller than 5 px and remove those regions which area is smaller than 5 px.

4.1.5. Refinement of region boundaries

During the slide preparation process a tissue specimen is sliced into sections approximately $4\ \mu\text{m}$ thick. Therefore, in void spaces we may often notice deeper tissue structures showing through the half-transparent paraffin wax used to fix the specimen (Fig. 4.7). This problem is severe, as usually void spaces between collagen fiber bundles cover over 30% of dermal area [81], while dermis itself constitutes about 90% of the whole histopathological skin specimen [78].



Figure 4.7. Deeper tissue structures showing through the half-transparent paraffin wax used to fix the specimen.

These problematical regions are dealt with in the background segmentation in the last stage of our method. To include them, we perform a hysteresis thresholding analogous to the one performed in (4.9). This time, however, we consider also the fact that due to deeper tissue layers showing through the paraffin wax, the average lightness of background regions confined within the tissue (i.e., holes) is typically lower than the average lightness of background regions outside the tissue.

The average background lightness of a given image pixel is calculated by applying an operation we called a “selective filtering”. It produces the resulting image using information from only those pixels in the input image which are included in the ROI mask as ones:

$$I_{\text{sf}} = (I * K - (I \cdot \overline{M}) * K) / (M * K) \quad (4.10)$$

where I is the input image, K is the filter kernel, and M is the ROI mask (both image multiplication $A \cdot B$ and division A/B are performed pixel-wise). We performed a Gaussian filtering with kernel $\sigma = 5.0$ (and kernel size 21 px) to the L^* channel of the blurred input image (i.e., the result of the preprocessing stage) using M_{Bg}^0 mask as the ROI mask.

Let us define the following condition:

$$\text{Cond}_3(c_P) = (L_P^* > L_{\text{BgAvg}}(P) - 1.32\text{JND}) \wedge (C_P^* < 3.06\text{JND}) \quad (4.11)$$

It is satisfied by pixels having both lightness similar to the average local slide background lightness L_{BgAvg} and low chroma.

We obtain the final tissue segmentation mask M_{Bg} by performing the following morphological reconstruction (the condition Cond_3 is applied pixel-wise for each pixel P):

$$M_{\text{Bg}} = \overline{\left(M_{\text{Bg}}^0 \xrightarrow{\text{8}} M_{\text{Bg}}^0 \cup \text{Cond}_3(c_P) \cap \left(M_{\text{Bg}}^0 \oplus \text{SE}_D(r) \right) \right)}. \quad (4.12)$$

The $X \cap \left(M_{\text{Bg}}^0 \oplus \text{SE}_D(r) \right)$ element restricts the reconstruction range to r pixels. We used $r = 5$ px as it was the maximum observed width of transition regions, regardless the image resolution.

4.2. Results

The presented method has been evaluated on large extracts from 60 skin WSIs from our four datasets: 10 from CMUJ-AM, 30 from CMUJ-SS, 10 from UMch, and 10 from UBC. To cover most of the variations which one would encounter in regular clinical practice, we chose images differing in staining quality, original magnification, and the capturing device (either an automatic microscope or a whole-slide scanner). The size of WSIs varied from 2000×1500 to 20000×30000 pixels. The ground truth was manually annotated by an expert dermatopathologist. We split the whole dataset into a training set consisting of 30 images (from each dataset we randomly selected 2 images) and a validation set consisting of 30 images.

4.2.1. Statistical analysis

We evaluated the performance of our method by measuring the overlap of the segmented tissue mask S to the manually labeled ground truth mask G using the Jaccard index, defined as:

$$J(S, G) = |S \cap G| / |S \cup G| \quad (4.13)$$

where $J(S, G) = 1$ represents a perfect overlap with the ground truth.

For each image we computed the Jaccard index in three manners: using the whole image region (denoted as J_{all}), with outer background region neglected (denoted as J_{in}), and when considering only the outer background regions (denoted as J_{out}). We consider J_{in} index more informative than J_{all} index, since it is somewhat simple to segment outer background regions, which oftentimes occupy over 50% of the image area (resulting in bloated values of Jaccard indexes).

We formulated the J_{in} and J_{out} indexes as follows:

$$J_{\text{in}}(S, G, M_{\text{Bg}}) = J(S \cap \overline{M_{\text{BgOut}}}, G \cap \overline{M_{\text{BgOut}}}) \quad (4.14)$$

$$J_{\text{out}}(S, G, M_{\text{Bg}}) = J(S \cap M_{\text{BgOut}}, G \cap M_{\text{BgOut}}) \quad (4.15)$$

where M_{BgOut} is the outer background region mask. The M_{BgOut} mask is obtained by performing the opening of M_{Bg} mask using a disk-shaped structuring element of radius 15 px and retaining only those regions which are adjacent to image edges.

A manual labeling of the whole high-resolution WSI up to single pixels would definitely be an immensely laborious and subjective task, leading to distorted results. Therefore, we discarded regions smaller than 30 px both in ground truth masks and in segmentation masks.

4.2.2. Parameter selection

To choose the proper values of chroma threshold τ_{Hist} and the parameter q for histogram construction (in Section 4.1.3) as well as the values of τ_{Core} and τ_{Base} thresholds in (4.6) and (4.7), respectively, the sequential quadratic programming (SQP) approach has been used. The cost function f_c has been defined as

$$f_c(\tau_{\text{Hist}}, q, \tau_{\text{Core}}, \tau_{\text{Base}}) = -\frac{1}{n} \sum_{i=1}^n (\bar{J}_{\text{in}}^i + \bar{J}_{\text{out}}^i) / 2 \quad (4.16)$$

where n is the number of datasets (here: 4), and \bar{J}_{in}^i and \bar{J}_{out}^i are the means of J_{in} and J_{out} computed for all training images from the i th dataset using the provided parameter values. The imposed constraints included $\tau_{\text{Hist}} \in \langle 2.0, 8.0 \rangle$, $q \in \langle 0\%, 50\% \rangle$, $\tau_{\text{Core}} \in \langle 0.5, 5.0 \rangle$ and $\tau_{\text{Base}} \in \langle 1.0, 12.0 \rangle$, whereas the initial point has been set to $(\tau_{\text{Hist}}^0, q^0, \tau_{\text{Core}}^0, \tau_{\text{Base}}^0) = (3.0, 1\%, 2.5, 6.0)$ and the minimum step size to 0.1. The training set included randomly selected half of images from each WSI dataset (i.e., 30 images in total).

The optimization algorithm found the best solution at $(\tau'_{\text{Hist}}, q', \tau'_{\text{Core}}, \tau'_{\text{Base}}) = (4.075, 0.5\%, 3.047, 7.049)$, which can be expressed using multiplicities of JND as $(1.77\text{JND}, 0.5\%, 1.32\text{JND}, 3.06\text{JND})$. Typical histograms of color distance distributions for pixels belonging to background and tissue regions, as well as the location of optimal threshold values, are shown in Figure 4.8.

Since training slides demonstrated substantial variability in both staining and illumination (as shown in Table 4.1), we expect that the adopted parameter values will allow to obtain satisfactory results also for most new cases (provided that they lack significant artifacts, like strong vignetting). To verify this assumption, boxplots of J_{all} , J_{out} and J_{in} indexes computed for training and test sets were compared

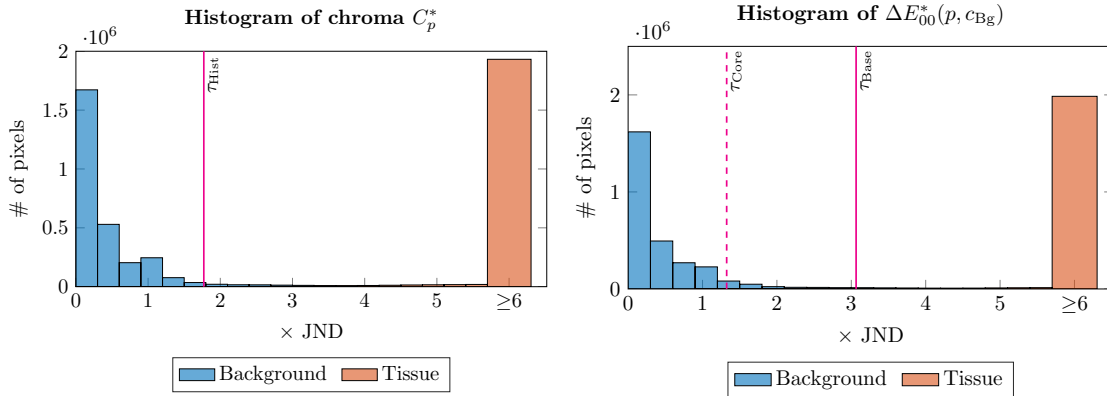


Figure 4.8. Typical histograms of chroma C_p^* and the distance to the “mean” background color $\Delta E_{00}^*(p, c_{Bg})$ constructed using data from the same WSI.

(Fig. 4.9): in the testing set there is only one serious outlier (for which $J_{all} = 0.64$ and $J_{in} = 0.19$), caused by the above-mentioned strong vignetting, whereas in all cases $J_{out} > 97\%$.

Table 4.1. The average CIELAB color of the outer slide background for three different slides

Dataset	CIELAB color	Color sample
CMUJ-AM	(87.30, 0.38, 0.53)	
CMUJ-SS	(90.85, -0.84, 0.49)	
UBC	(96.01, 0.00, 0.00)	

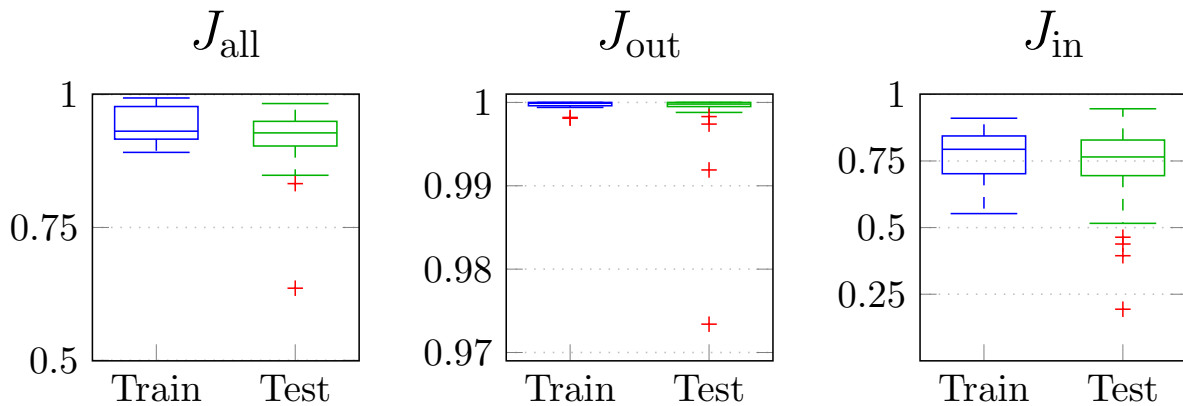


Figure 4.9. Visualization of Jaccard indexes as boxplots for the training and the testing datasets (outliers are marked with a “+” mark)

4.2.3. Experimental results

For 35 out of 60 images the proposed algorithm matched the actual tissue regions without notable differences and scored area overlaps J_{all} and J_{in} of more than 90% and 75%, respectively. For the remaining 25 images the problematic issue was the detection of tiny blank spaces (mainly between bundles of collagen fibers) which resulted in slightly oversegmented tissue regions (Fig. 4.11). However, in each investigated case the algorithm included all relevant parts of the tissue and even when slide preparation artifacts

were present yielded results acceptable for a general-purpose method (i.e., for all cases $J_{\text{out}} > 97\%$). In worst case the area overlaps J_{all} and J_{in} equaled 82% and 52%, respectively. Such unsatisfactory performance resulted from significant vignetting artifacts shown in Figure 4.10. With a deviation of 5.3% in the J_{all} overlap, the proposed method yields a consistent foreground/background separation. The J_{out} mean was 99.91%.

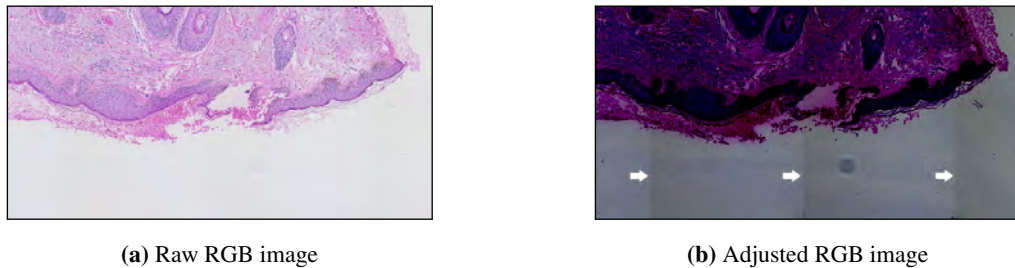


Figure 4.10. The vignetting artifacts: (a) the raw RGB image, and (b) the same image with the colormap adjusted to emphasize artifacts (i.e., for each channel its intensity values from the range $\langle 200, 255 \rangle$ were mapped to the range $\langle 0, 255 \rangle$).

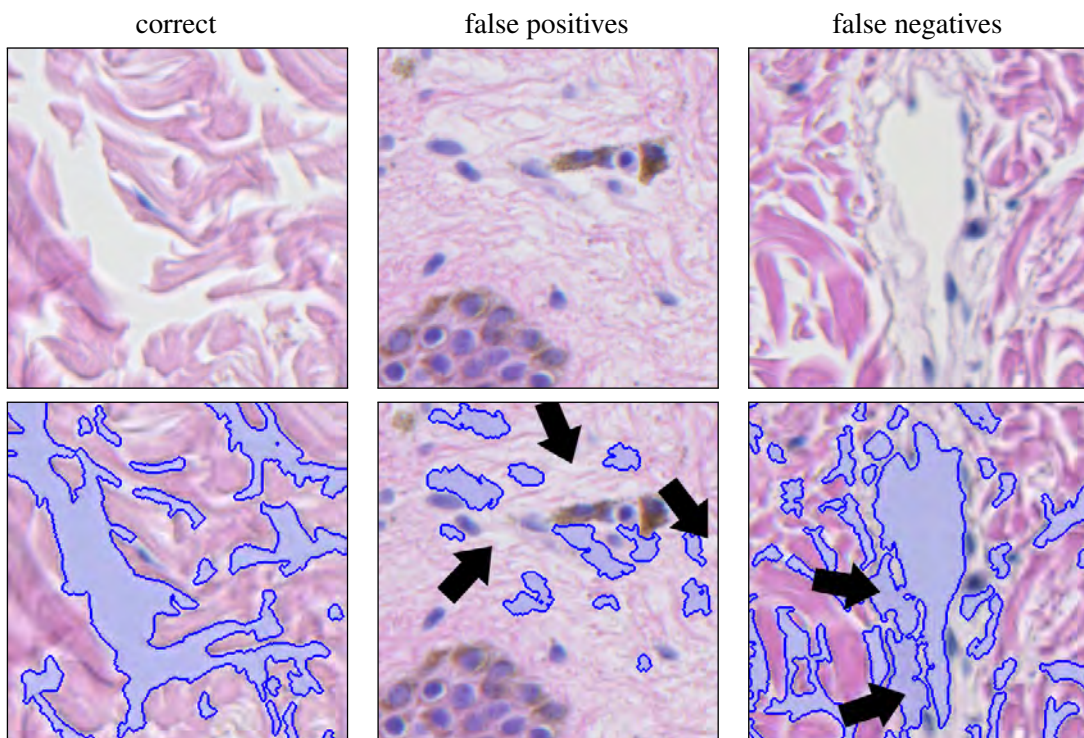


Figure 4.11. Qualitative results yielded by the proposed algorithm. Both cases presents close-ups of the same collagen section (200×200 px) with detected background marked blue. Large- and medium-sized void spaces between bundles of collagen fibers are accurately segmented, while some tight spaces are erroneously detected as the tissue. The proposed algorithm has a tendency to include very pale tissue structures (e.g., blood vessels) into the background region.

We compared the segmentation results of our method with the results obtained using the Foreground Extraction from Structure Information (FESI) and GrabCut (GC) methods [106, 108].

Results for FESI method were obtained using the Python implementation provided by Bug et al. [106]. Since the resolution images in our dataset differed from those used by Bug et al. (they analyzed the 5th resolution level of images instead of the 1st level), to find the optimal set of parameters for our dataset a grid-search procedure with the following ranges of parameter values was conducted: $k_g = 5\text{--}15$ px (step: 2), $\sigma = 2\text{--}4$ (step: 1), $k_m = 15\text{--}45$ px (step: 5), $n = 5$, $r = 3\text{--}7$ px (step: 2), and $d = 10\text{--}100$ px (step: 30). For the training dataset best results were obtained using the following combination of parameter values: $k_g = 15$ px, $\sigma = 4$, $k_m = 15$ px, $n = 5$, $r = 7$ px, and $d = 10$ px.

Results for GrabCut method were obtained using the implementation provided by OpenCV 3.4 (see the `cv.grabCut()` function). To increase their precision, a ROI rectangle 20 px from the edges of the WSI was defined in order to include some hard-labeled background pixels in the initial estimate and the algorithm was run for 15 iterations.

In Table 4.2 and Figure 4.12 we presented the summary statistics of the Jaccard index across all 60 images. Some qualitative segmentation results are presented in Figure 4.13.

Table 4.2. Mean and standard deviation of Jaccard indexes J_{all} and J_{in} for GC, FESI and the proposed method. All three methods achieved the mean J_{out} of 0.999–1.000 with standard deviation of 0.001.

Method	J_{all}		J_{in}	
	μ_J	σ_J	μ_J	σ_J
GC	0.776	0.112	0.243	0.174
FESI	0.695	0.154	0.243	0.132
proposed	0.929	0.053	0.749	0.140

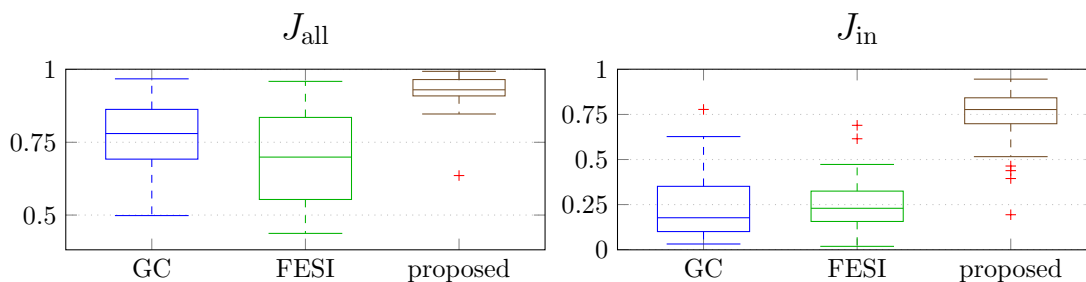


Figure 4.12. Visualization of Jaccard indexes as boxplots (outliers are marked with a “+” mark)

On average our approach provides better results than the traditional GC and FESI methods: J_{all} mean equaled 0.929 vs. 0.776 and 0.695. The results of our method are also more stable: J_{all} standard deviation of 0.053 vs. 0.112 and 0.154. What is more important, our approach detects void spaces within tissue substantially better than other methods (J_{in} mean of 0.749 vs. 0.243 and 0.243).

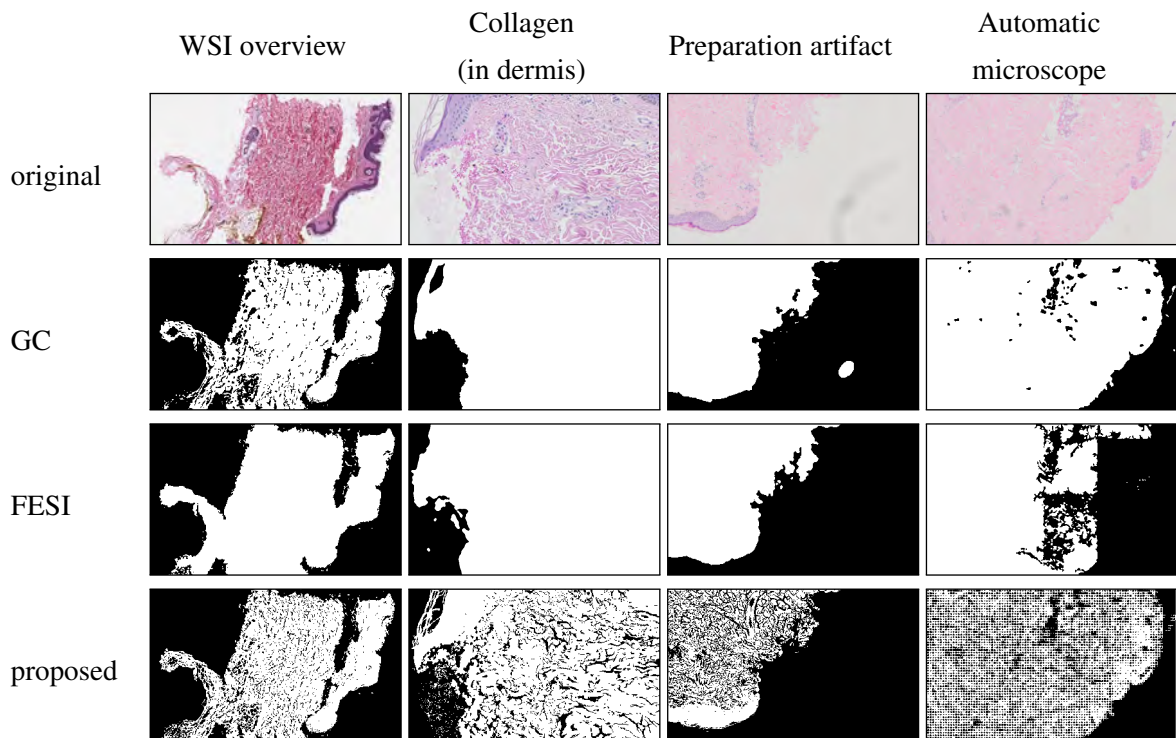


Figure 4.13. Qualitative results for the GC, FESI, and the proposed algorithm. Column 1: A properly stained WSI captured using a whole slide scanner. All the methods correctly detected outer tissue boundaries. Column 2: A close-up on a collagen section. Only the proposed method is able to properly segment the small-grained void spaces between bundles of collagen fibers. The GC and FESI detects only larger pores. Column 3: Both FESI and the proposed method can identify some slide artifacts (such as dust and hair under the cover slip) as background. Column 4: A challenging WSI exhibiting weak staining and captured using an automatic microscope (resulting in an inferior image quality when compared to the ones taken using slide scanners). FESI fails to cope with the “tiled” image structure produced by an automatic microscope.

4.3. Conclusions and discussion

The proposed method yielded satisfactory results for skin WSIs from all four datasets, regardless of staining quality and resolution of images. It segments the tissue (i.e., foreground) accurately and consistently, with a total area overlap of more than 92.9% and a standard deviation of 5.3% in the overlap. However, at the same time it has a tendency to undersegment areas on edges of collagen fibers bundles and the void spaces between tightly packed collagen bundles (Figures 4.11 and 4.13).

Both GC and FESI methods has severe problems to properly detect the tissue “porosity”, i.e., void spaces within the tissue structure, as they typically detect only large pores and in certain cases even totally ignore porosity (columns 1 and 2 in Figure 4.13). It was an expected result above all for the FESI method, as it was designed to segment tissue at the 5th layer of multi-resolution WSIs when the pixel spacing is approximately 7–8 $\mu\text{m}/\text{px}$. Since the ratio of void spaces within a skin specimen to the total area of background regions may amount to 30–40% and the spacing between bundles form

thin “ribbons” with average width of as little as 2–5 μm (i.e., 8–20 px in $\times 40$ images captured using CCD sensors which resolution is 0.25 $\mu\text{m}/\text{px}$) [81], neglecting the subtle tissue “porosity” renders both GC and FESI methods unfeasible for practical application in numerous image processing workflows for digital dermatopathology.

The methods proposed by Bándi et al. [107] were designed specially to segment the 4th layer of WSIs where the pixel spacing was 3.84 $\mu\text{m}/\text{px}$. However, the spacing between collagen bundles in such low resolution is only about 1 px wide, thus virtually preventing correct segmentation of the void spaces between such bundles. That problem was even mentioned and demonstrated in qualitative results published in [107]. Consequently, we did not compare our method to the methods by Bándi et al.

4.4. Summary

In this chapter we presented an algorithm for tissue segmentation in high-resolution WSIs of H&E-stained skin specimens, based on color statistical analysis, CIELAB color thresholding, and binary morphology. We provided the segmentation results and compared them quantitatively with other methods.

5. Epidermis segmentation

The morphometric and cytologic features of the epidermal area are among most important factors considered when diagnosing skin conditions and during skin tissue grading [16, 109]. Even though a few computerized epidermis segmentation methods have already been proposed, in our opinion this topic has not been sufficiently explored yet. The methods already described in the literature are based mainly on the combination of global thresholding technique with area and shape analysis, making substantial assumptions about color and contrast of WSI images. However, inter- and intra-image variations in staining and illumination are quite common and stain distribution will be strongly affected by the presence of certain pathological changes (as mentioned in sections 2.2.3 and 2.5). Consequently, these assumptions are often not met when it comes to analyzing data from different sources.

In this chapter we propose a novel approach towards epidermis segmentation, based on our method published in [24]. In the initial segmentation stage shape and distribution of void regions in the tissue are analyzed. Then, it is refined by analyzing the concentrations of hematoxylin and eosin stains in the tissue. In the final stage, shape and location constraints are applied on the segmented region to obtain the actual epidermal region. Since our method strongly relies on domain-specific knowledge of cell and tissue morphology and their biochemical properties, it is robust to differences in staining protocols, stain absorption within individual images, and to illumination variability. We provided the segmentation results, and compared them both quantitatively and qualitatively with other methods for epidermis segmentation.

5.1. Algorithm for automatic epidermis segmentation

The presented epidermis segmentation algorithm combines the analysis of shape and distribution of void regions in the tissue specimen and the analysis of H&E stain concentrations across the image. It heavily utilizes the domain-specific knowledge of both morphometric and biochemical properties of skin tissue structures (such as their size, frequency of occurrence, color, and way of staining); values of those parameters and constants used in our method which are based exclusively on measurements of morphometry of cells and tissue structures found in the human skin are summarized in Section 2.4.5. Figure 5.1 shows the schema of the proposed segmentation pipeline.

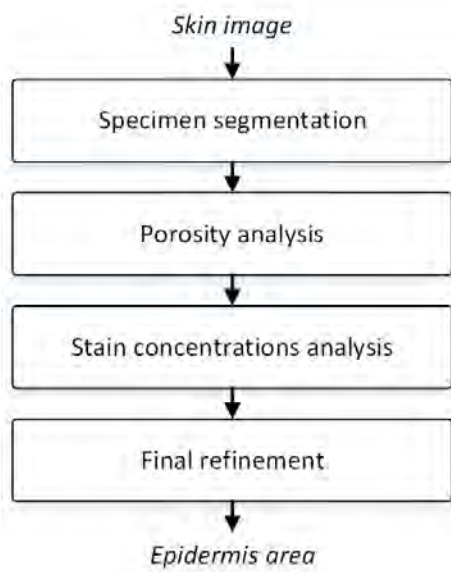


Figure 5.1. Schema of the proposed segmentation pipeline.

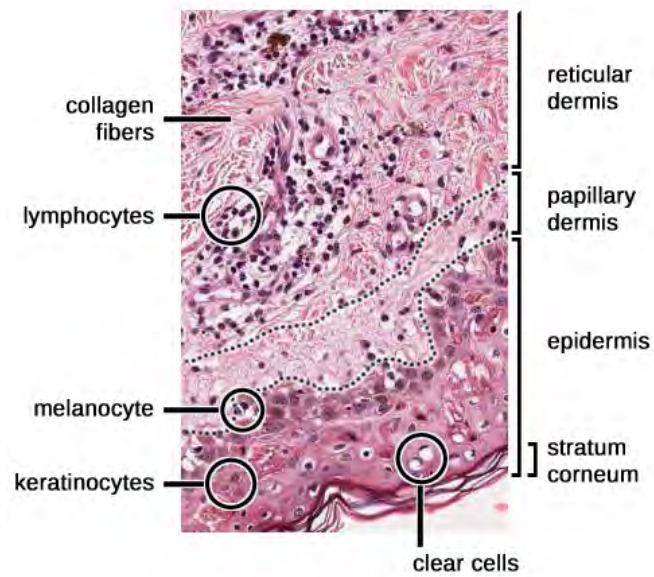


Figure 5.2. Structures of interest marked on a H&E-stained skin section.

5.1.1. Specimen segmentation

The preliminary stage of our method is specimen segmentation. To obtain the initial foreground (tissue) mask M_{Fg} we utilize the method described in Section 4.1. The background mask M_{Bg} is simply the negation of the foreground mask: $M_{\text{Bg}} = \overline{M_{\text{Fg}}}$.

Additionally, we compute the M_{BgOut} mask representing the “outer” slide background (i.e., background regions not confined within tissue) by retaining only the largest region in the $M_{\text{Bg}} \oplus \text{SE}_D(8\mu\text{m}/R_I)$ mask (as typically void spaces between collagen form ribbons which are at most $15\mu\text{m}$ wide).

5.1.2. Porosity analysis

Porosity is a measure of the void spaces (pores) in a material, and is a fraction of the volume of voids over the total volume. In case of histological skin specimens, such spaces are found mostly in the following four regions: in fat tissue, between bundles of collagen fibers, near cell nuclei, and on intercellular junctions. Void spaces are the result of tissue processing procedure (shrinkage effects of formalin fixation, paraffin embedding) and are considered artifacts. The bulk of void spaces is located in fat tissue layer and between dermal bundles of collagen fibers. They correspond to fat and the ground matrix, respectively, which are both lost during the tissue processing procedure. On the other hand, void spaces near cell nuclei are in fact poorly stained cytoplasm and gaps between cells correspond to half-opaque desmosomes.

In contrary to dermis and adipose tissue, epidermis contains hardly any transparent regions. Most of them are either clear cells or desmosomes, hence their shape is well-defined and differs from the shape

of void spaces in dermis and in fat tissue allowing for the coarse segmentation of epidermis by analyzing the geometry and distribution of pores in the lesion.

In this stage we firstly identify those void spaces which are most likely clear cells or desmosomes and include them into the M_{Fg} mask. The detection of every single transparent region is not the point, the aim is rather to flood-fill the bulk of large holes and enough smaller ones, so that porosity of epidermis will differ significantly from porosity of deeper skin layers. Then we exclude image regions with sufficiently high porosity, as they are most likely part of deeper skin layers (dermis, adipose tissue). Figure 5.3 shows the schema of this segmentation stage.

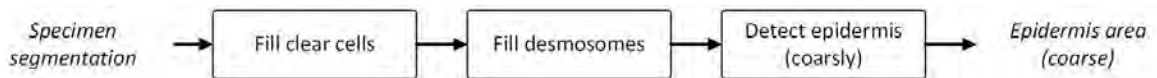


Figure 5.3. Schema of the porosity segmentation stage.

The algorithm described in this section uses a procedure we called “computing tissue density”. For each image pixel this procedure computes the share of ones in this part of pixel’s neighborhood which belongs to the given mask M (the density of pixels which do not belong to M is set to zero). The tissue density ρ at pixel P is defined as

$$\rho(P) = \begin{cases} \frac{T(P)}{|SE|} & \text{if } P \in M \\ 0 & \text{otherwise} \end{cases} \quad (5.1)$$

where SE is a binary structuring element defining the neighborhood, and $T = M * SE$ is a matrix containing the number of neighboring pixels belonging to the given mask M for all image pixels in M (obtained using the convolution).

Filling the clear cells

Clear cells have abundant glycogen causing their cytoplasm to stain poorly in routine H&E staining (Fig. 5.4). Clear cells found in sections of light skin may be of one of the following types: melanocytes, Langerhans cells, or Merkel’s cells. All these three types of cells share similar geometric characteristics and are located in epidermis [28]. We identify clear cells by assessing the shape of each individual void region (i.e., each 8-connected component) in the M_{Bg} , which should be convex and elliptic.

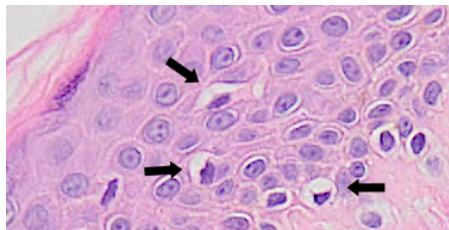


Figure 5.4. Examples of clear cells found in epidermis.

The initial algorithm verifying whether a given transparent region $CC_{0,i}$ is a valid clear cell candidate consists of the follows steps:

- (a) Reject the $CC_{0,i}$ if $A(CC_{0,i}) > A_{\text{KerMax}}$ (the diameter of a Langerhans cell is comparable to the diameter of a keratinocyte, melanocytes are smaller than basal keratocytes [122]).
- (b) Flood-fill all holes (so that nuclei and pixels not considered transparent due to intense noise or larger dust particles are also included).
- (c) Reject tiny spurs by applying the morphological opening with a disk-shaped structuring element of radius $r = 1$.
- (d) Apply the morphological thickening by 1 pixel to restore cell's core pixels lost in step (c).
- (e) Apply the bit-wise AND operation on masks obtained in steps (d) and (b) to exclude pixels exceeding outer boundaries of the (initial) $CC_{0,i}$ region.

The change of candidate region $CC_{0,i}$ after each step is shown in Figure 5.5.

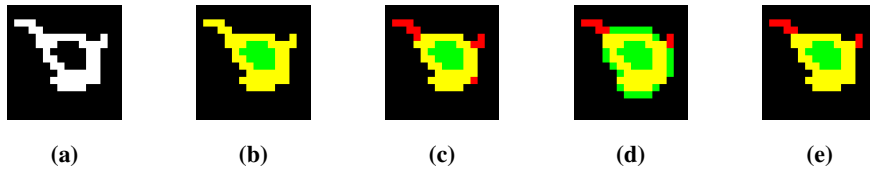


Figure 5.5. Changes to the candidate region after each step of the clear cell candidate verification algorithm, with respect to the initial region $CC_{0,i}$ (red – pixels removed, green – pixels added).

Let us denote the region obtained as the result of step (e) as CC_i . If the area of this region is less than $0.7A(CC_{0,i})$ we reject such a region, as components with many spurs are most likely void spaces between collagen fibers. To evaluate the criterion related to convexity and ellipticity of the candidate region CC_i , we analyze its eccentricity and solidity. The eccentricity is the ratio of the distance between the foci of the ellipse that has the same second-moments as the region to the length of its major axis. Its values fall between the range of 0 and 1 (both 0 and 1 are degenerate cases: the first is actually a circle, while the second is a line segment). The solidity is the ratio of region area to the area of its convex hull (a convex region has solidity of 1).

To verify whether a clear cell candidate region CC_i is a valid candidate or not, we put the following constraint:

$$CC_i = \begin{cases} \text{valid candidate} & \text{if } 5 \leq A(CC_i) \leq A_{\text{KerMax}} \wedge \text{Solidity}(CC_i) > 0.7 \\ & \wedge \text{Eccentricity}(CC_i) < 0.95 \\ \text{invalid candidate} & \text{otherwise} \end{cases} . \quad (5.2)$$

An ideal cell candidate would be a circle (which is a convex shape), thus having $\text{Eccentricity}(CC_i) = 0$ and $\text{Solidity}(CC_i) = 1$. However, the clear cell nucleus is often pushed towards the cell's boundary resulting in the transparent region of the clear cytoplasm which resembles a crescent (a slightly concave, elongated shape). Consequently, we relaxed the criteria for both eccentricity and solidity.

In the last step we classify a given initial (not morphologically processed) transparent region $CC_{0,i}$ as either a clear cell or a non-clear cell based on its geometry. Valid unprocessed candidates are obtained by reconstructing a mask containing valid processed candidate regions into the M_{Bg} mask. Then the

classification of each $CC_{0,i}$ region is conducted using the following criterion:

$$CC_{0,i} = \begin{cases} \text{clear cell} & \text{if } a_{maj}(CC_{0,i}) \leq 2r_{\text{KerMax}} + 1 \wedge \text{Solidity}(CC_{0,i}) > 0.5 \\ & \wedge \text{Eccentricity}(CC_{0,i}) < 0.95 \\ \text{non-clear cell} & \text{otherwise} \end{cases}, \quad (5.3)$$

where $a_{maj}(CC_{0,i})$ is the major axis length of the ellipse having the same normalized second central moments as the $CC_{0,i}$ region (it must not exceed the maximum diameter of a keratinocyte). The solidity threshold is different than in (5.2), as a raw region $CC_{0,i}$ often contains small spurs. Regions recognized as clear cells (or rather: their cytoplasm) are included into the M_{Fg} mask. Threshold values in both (5.2) and (5.3) were chosen empirically, based on the procedure described in Section 5.2.1.

Filling the desmosomes

Desmosomes are intercellular junctions of epithelia, particularly abundant in tissues continually assailed by mechanical forces (such as epidermis and myocardium) [123]. They are less than $1 \mu\text{m}$ in diameter and normally appear as half-transparent (Fig. 5.6). However, as the result of poor staining parts of desmosomes are often detected as void spaces.

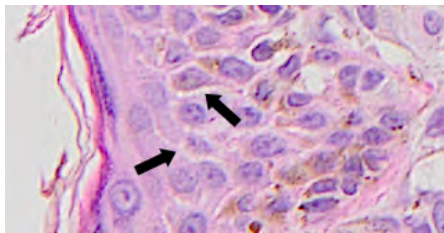


Figure 5.6. Desmosomes encompass the intercellular space between adjacent keratinocytes.

Candidates for desmosomes are void regions forming lines. Since we analyzed images at about $0.5 \mu\text{m}/\text{px}$, each such a line would be 1 px thick. We perform the line detection by counting the number of invalid combination of ones in an object using the hit-or-miss operation (the invalid combinations are shown in Figure 5.7). An ideal line would contain only valid combinations and no loops. However, the image registration process is hardly ever perfect (e.g., due to the CCD noise) and artifacts must be taken into account. To deal with them, we introduce an *error tolerance* parameter, which describes the maximum number of invalid combinations in an object when it is still acceptable to consider it a line (we set the error tolerance to 1).

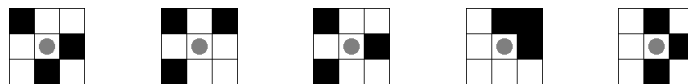


Figure 5.7. Combinations of ones resulting in the object being marked as an “invalid line” (the central pixel of the hit-or-miss structuring element is marked with a gray circle). Each of these combinations rotated by 90° , 180° , and 270° is also considered invalid.

Not only (epidermal) desmosomes may appear as a line of transparent pixels, narrow void spaces between bundles of collagen fibers in dermis often have similar appearance. To distinguish between them we use the following criteria: (1) Since desmosomes are half-opaque, it is seldom the case that

they appear as transparent on its whole length. Therefore, even though desmosomes run along the whole perimeter of every cell, the maximum length of a transparent section of a desmosome is set to half the perimeter of a largest observed keratinocyte ($L_{\max} = 2\pi r_{\text{KerMax}}$). (2) The desmosome candidate region should at least partially overlap with a region of high tissue density (the density is computed in the disk-shaped neighborhood of radius $r_{\text{EpidermMin}}$). The threshold value for maximum density is set to $\rho_{\text{Desm}} = 0.85$, as collagen fibers constitute roughly 70% of the weight of dry dermis while void spaces between bundles of collagen fibers in papillary dermis are smaller than in reticular dermis [70]. Candidate lines satisfying both the aforementioned criteria are included into the M_{Fg} mask.

Detecting epidermis

Now, after filling most tiny and medium-sized pores in epidermal region, the remaining ones are sparse and located primarily in outermost layers of epidermis. On the contrary, porosity of the dermal region (particularly of the reticular dermis) remained high, since void spaces between collagen fibers bundles were left unfilled.

Therefore, we firstly compute the “epidermal core” region by thresholding the tissue density map of the M_{Fg} mask computed for radius $r_{\text{EpidermMin}}$ with the value $\rho_{\text{EpidermMin}}$. As epidermis is the outermost skin layer, we retain only those chunks of core regions which lie in the proximity of lesion’s boundary. The proximity is set to $d = d_{\text{SCAvg}} + r_{\text{EpidermMin}}$ (the addition of the $r_{\text{EpidermMin}}$ factor was introduced to handle the corner case in which a core chunk is actually an isolated pixel placed in the center of the otherwise empty neighborhood of radius $r_{\text{EpidermMin}}$). To restore the actual thickness of epidermal region we dilate the retained core chunks with a disk-shaped structuring element of radius $r_{\text{EpidermMin}}$. Then, we perform a binary AND operation with the M_{Fg} mask to remove excessive pixels. Lastly, we further reduce the share of false positive pixels by rejecting regions with area smaller than the area of 80 maximal keratinocytes ($80A_{\text{KerMax}} \approx 0.035 \text{ mm}^2$). The coarse epidermis segmentation mask obtained in this stage will be further denoted as M_{SegPor} .

5.1.3. Stain concentrations analysis

In the course of the porosity analysis stage large regions of dermal collagen and fat tissue were rejected from the epidermis segmentation mask. However, it usually still includes prominent false positive regions, like chunks of tightly packed collagen fibers in papillary dermis, nests of melanocytes, or thicker sections of stratum corneum. To address this issue we analyze concentrations of both hematoxylin and eosin stains in each pixel. Since both collagen and corneocytes (cells forming stratum corneum) are eosinophilic [124], which means they absorb much eosin and little hematoxylin, these problematic tissue structures may be filtered out using criteria based on a combination of thresholding of both individual stain concentration maps and maps of hematoxylin-to-eosin concentrations ratio. As soon as collagen regions are excluded from the segmentation mask, nevi cell and lymphocytic infiltration may also be removed by rejecting regions located deep inside the lesion (the procedure is similar to the one described in Section 5.1.2). Figure 5.8 shows the schema of this segmentation stage.

As both collagen and corneocytes lack nuclei [125], throughout this section we often formulate criteria which include the $C_{\text{NucMin}}^{\text{H}}$ parameter representing the minimal normalized concentration of hematoxylin in a region of a typical cell nucleus.

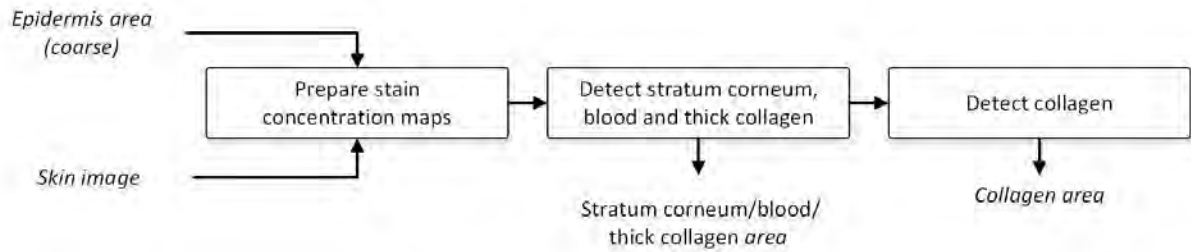


Figure 5.8. Schema of the stain concentrations analysis stage.

Preparing stain concentration maps

The raw stain concentration maps are obtained using the stain unmixing method proposed by Macenko et al. [58], as implemented by Mitko Veta [126]. This method assumes there exists a specific stain vector corresponding to each of the two stains (hematoxylin and eosin) present in the image, and that in optical density space the actual color of every pixel is simply a linear combination of these stain vectors. As the weight of each component must be non-negative, all values must fall between the two stain vectors. For each individual WSI image the algorithm by Macenko et al. estimates those particular stain vectors and uses these estimates to compute stain concentration maps.

There are two key parameters of the above-mentioned stain separation algorithm. The first one is α : α^{th} and $(100 - \alpha)^{\text{th}}$ percentiles are used to robustly determine minimum and maximum stain concentrations in an input image. We set it to the value proposed in the original paper, i.e., $\alpha = 1$. The second one is a threshold value β governing the thresholding of optical density map used to determine slide background region (background regions are excluded from the process of estimating the stain vectors thus improving method's stability). Instead of using the β parameter for thresholding in optical density space, we used the M_{Fg} mask computed in Section 5.1.1. We also excluded from analysis those slide regions where black ink used to mark the surgical margins was present, as it would negatively affect the stain vector estimation process [27] (ink segmentation has been performed using our method described in [27]).

To preserve the visual contrast between the hematoxylin and eosin stains in their concentration maps, both raw maps are normalized to reference maximum stain concentrations, $C_{\text{RefMax}}^{\text{H}}$ and $C_{\text{RefMax}}^{\text{E}}$, using values provided in [126]. Since the stain unmixing algorithm produces noisy results, we smooth normalized maps by applying the selective smoothing procedure, which averages values only over pixels belonging to the (foreground) mask M thus preventing artifacts on the edge of regions in M . The procedure uses a standard Gaussian filter (the kernel size k is set according to the formula $k = 2\lceil 2\sigma \rceil + 1$). The selective smoothing of image I with a filter kernel h over the foreground mask M requires the calculation of two auxiliary images: $S_{\text{Fg}} = (I \cdot M) * h$ representing the proportion of foreground pixel values in the smoothed image, and $W_{\text{Fg}} = M * h$ representing the sum of foreground pixel weights. Pixel P in the

smoothed image I_σ is then computed according to the formula

$$I_\sigma(P) = \frac{S_{\text{Fg}}(P)}{W_{\text{Fg}}(P)} \quad (5.4)$$

In our method we set $\sigma = (r_{\text{KerAvg}} - r_{\text{KerNuc}})/R_I$ (to “fill” the region of cytoplasm between nuclei of keratinocytes with hematoxylin, ensuring a sufficiently high average level of hematoxylin across the whole epidermis) and used the M_{SegPor} mask as the smoothing foreground mask.

Detecting blood and stratum corneum

In this step we reject blood hemorrhage (which does not constitute epidermis) and parts of stratum corneum (which is not relevant for the diagnosis of most skin diseases). Both of these tissue structures are strongly eosinophilic and, as they have only a minimal amount of basophilic components, at the same time they absorb little hematoxylin [124]. Therefore, to classify an image pixel P as either blood or a part of stratum corneum we use the following criterion based on stain concentrations:

$$M_{\text{Red}}(P) = \begin{cases} 1, & \text{if } C^{\text{E}}(P)/C^{\text{H}}(P) > 4C_{\text{RefMax}}^{\text{E}}/C_{\text{RefMax}}^{\text{H}} \wedge C^{\text{H}}(P) < C^{\text{NucMin}} \text{H} \\ & \wedge C^{\text{E}}(P) > 0.65C_{\text{RefMax}}^{\text{E}} \\ 0, & \text{otherwise} \end{cases}, \quad (5.5)$$

where $C^{\text{H}}(P)$ and $C^{\text{E}}(P)$ are concentration of hematoxylin and eosin, respectively, in pixel P . Threshold values in (5.5) were chosen empirically, based on the procedure described in Section 5.2.1.

Detecting collagen

At this stage of the segmentation pipeline the current epidermis segmentation mask includes mainly two elements: epidermis and collagen. Since collagen fibers in papillary layer has a fine structure, the eosin concentration in its corresponding image regions is much lower than in the regions where thick collagen fibers bundles are present (e.g., in the reticular layer) [124]. Moreover, there are numerous basophilic elements interleaved with collagen, such as fibroblasts, lymphocytes, and nevi cells. Consequently, the eosin-to-hematoxylin ratio used to segment blood and stratum corneum is an unsatisfactory criterion basis for papillary collagen rejection.

To identify papillary collagen regions we analyze the two dimensional (2D) histogram of pixel counts with respect to eosin and hematoxylin concentrations. The histogram is constructed only from tissue pixels located in the proximity of the outer border of the tissue specimen (i.e., from the M_{BgOut} mask). Typically, there are two significant peaks on such a histogram: one corresponding to collagen and another corresponding to basophilic structures (and their neighborhood). We identify these peaks, determine their “zones of influence” and exclude pixels forming the “collagen zone” from the epidermis segmentation. The exact procedure for collagen detection is described below. We use the “Col” and “Oth” indexes in the notation of quantities to denote collagen and other (basophilic) tissue structures (keratinocytes, nevi cells, lymphocytes etc.), respectively.

Firstly, a normalized 2D histogram of H&E stain concentrations maps is computed for the tissue outer boundary zone $MBnd$ (it will be later denoted as H_0). We set the width of the zone to 250 μm

to maintain the balance between the amount of epidermis and collagen within that zone (the minimum thickness of cellular epidermis and stratum corneum is 40 μm and 30 μm , respectively). Each axis of the histogram is split uniformly into 75 bins, covering the range of $[0, C_{\text{RefMax}}^{\text{H}}]$ and $[0, C_{\text{RefMax}}^{\text{E}}]$, respectively. The objective when choosing the number of bins was to minimize the visual difference between image pixels in adjacent bins, to avoid quantization error (i.e., to obtain a continuous surface of the histogram), and to minimize the impact of CCD noise and JPEG artifacts on the results.

We identify the significant peaks in the 2D histogram (Fig. 5.9b) using the following procedure: In the preprocessing step, we create a duplicate of H_0 histogram, denoted as H' . In H' histogram bins for which the (normalized) hematoxylin concentration is over 95% or under 5% as well as those for which the (normalized) eosin concentration is over 95% or under 10% are suppressed – those bins are treated as being empty. The lower bound for eosin is greater than for hematoxylin to avoid searching for local maxima in the histogram region corresponding mainly to nests of nevi cells. Then, we repeat the following procedure until either enough regional maxima are found or the number of iterations exceeds the given threshold:

1. Set the initial number of quantization levels n to 5.
2. Quantize H' histogram into n levels using the following percentiles of 2D histogram: 100 – 2*n*, 100 – 2*n* + 2, 100 – 2*n* + 4, ..., 98, 100.
3. Search for 8-connected regional maxima in the quantized histogram and assess if two of them could correspond to the collagen and “other” peaks:
 - (a) Compute the centroid of each regional maximum. Choose the regional maximum with the centroid having the largest hematoxylin coordinate as the one corresponding to basophilic structures.
 - (b) From the remaining regional maxima choose the one encompassing the largest number of pixels and choose as the one corresponding to the collagen.
 - (c) Compute the weighted centroid for each of the above-mentioned regional maxima using the number of pixels in each bin as its weight. These centroids will be denoted as \bar{c}_{Col} and \bar{c}_{Oth} .
 - (d) The centroids \bar{c}_{Col} and \bar{c}_{Oth} must meet the following criteria: $\|\bar{c}_{\text{Col}} - \bar{c}_{\text{Oth}}\| \geq 0.1$ and the $\overrightarrow{\bar{c}_{\text{Oth}} - \bar{c}_{\text{Col}}}$ vector should have the slope of at least 60°. The first criterion helps to identify situations, when tiny fluctuations in the correctly segmented epidermis lead to the detection of two peaks. The second one reflects the fact that the basophilic structures forming the “other” peak must contain sufficiently more hematoxylin than the collagen (the “other” peak may not be located above the collagen peak).
4. If either insufficient number of regional maxima was found or they do not fulfill the requirements, increase the number of quantization levels n by 1 and repeat steps 2–4 (providing $n \leq 50$).

In some (rare) cases, due to the quantization, non-prominent local maxima are flattened. Therefore, if the procedure described above fails, regional maxima are identified in the H' histogram.

After enough regional maxima are identified and their location meets the constraints, we once again choose the collagen peak and the “other” peak – this time more precisely. The procedure is as follows:

1. Take two regional maxima encompassing the largest number of pixels and compute their centroids. Choose the maximum for which its centroid has higher hematoxylin coordinate and denote it as Max_1 .
2. If there were only two regional maxima identified (in total), assume the other one is Max_2 . Otherwise, from all regional maxima choose those having centroid with both higher hematoxylin coordinate and at the same time more pixels than Max_1 . If no such maxima were are, relax the condition on the hematoxylin coordinate – it is now enough if the difference in this coordinate is at most 2. If still no such maxima are found, drop the constraint concerning the number of pixels. From the maximum/maxima actually found choose the one encompassing the largest number of pixels and denote it as Max_2 .
3. Choose the regional maximum with the centroid having the largest hematoxylin coordinate as the one corresponding to basophilic structures. The other one corresponds to the collagen.

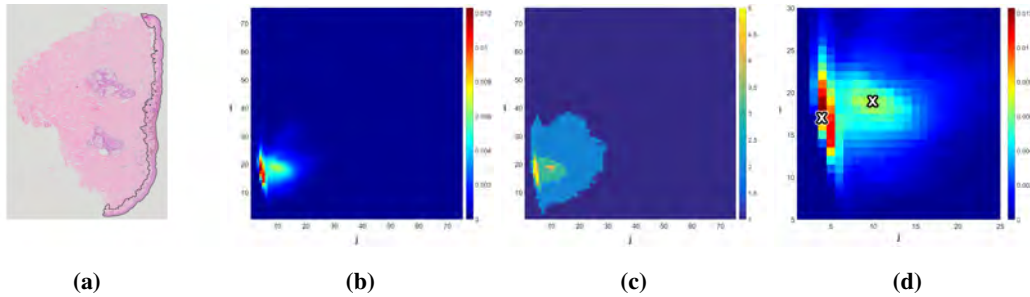


Figure 5.9. The peak detection procedure: (a) the registered RGB image I with M_{SegPor} mask marked black, (b) a 2D unquantized stain concentrations histogram of $I(M_{\text{SegPor}})$, (c) a 2D quantized stain concentrations histogram of $I(M_{\text{SegPor}})$ for $n = 5$, and (d) a region of interest of the unquantized histogram with weighted centroids marked with “X” marks. .

To delimit “zones of influence” of both peaks in the histogram (corresponding to collagen and to basophilic structures), we first identify the saddle point between these two peaks. The rationale for this step is as follows: In histological images concentration of both hematoxylin and eosin stain forms a smooth gradient, therefore the distribution of stain concentration values on a 2D histogram resembles a Gaussian mixture model with two significant components (one for basophilic structures and their proximity, and the other for collagen). The saddle point between peaks of those components is a bin for which the likelihood of an event that a pixel belonging to this bin is part of collagen equals the likelihood of an event that the pixel is part of other basophilic tissue structures. To identify the saddle point, the H' histogram is represented as an undirected weighted graph $G(V, E)$: its vertexes V correspond to histogram bins, its edges E connect vertexes corresponding to 4-connected bins, and each weight w_i of an edge $e_i = (v_1, v_2)$ is set to $w_i = |N(v_1) - N(v_2)|$ (where $N(v_j)$ is the number of pixels comprised within the v_j bin) (Fig. 5.10b). The saddle point v_{sp} between the collagen peak v_{Col} and the “other” peak v_{Oth} is defined as

$$v_{\text{sp}} = \arg \min_{v \in \pi_{\text{Min}}} N(v), \quad (5.6)$$

where π_{Min} is the shortest path between v_{Col} and v_{Oth} in G . The π_{Min} path is determined using Dijkstra’s algorithm [127]. Figure 5.10 shows the process of saddle point identification.

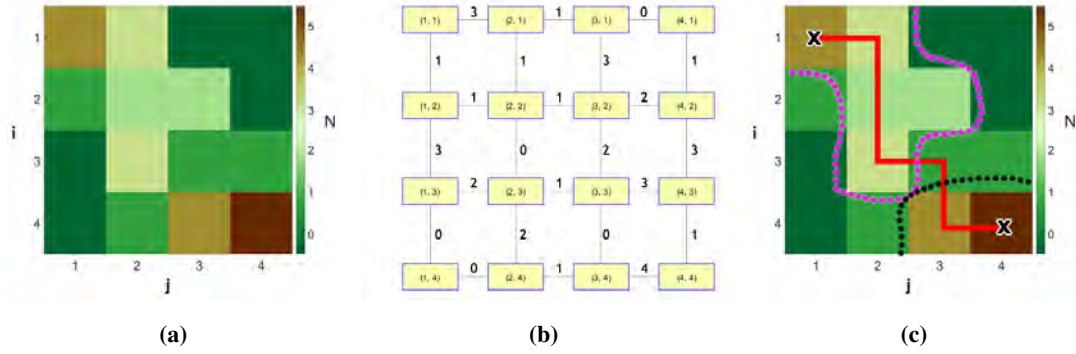


Figure 5.10. The saddle point identification process: (a) a 2D histogram H , (b) a graph representation $G(V,E)$ of the histogram H , (c) a ridge in the histogram H connecting two peaks (the red solid line) with saddle point at its minimum „height”.

The next step is to coarsely delimit “zones of influence” of both peaks in the H' histogram using the watershed transformation. To suppress insignificant local maxima in H' , we firstly use the following procedure to find such a lowest value $t_v > v_{sp}$ which, after being used as a threshold, fully separates the two peaks:

1. Set $t_v = v_{sp}$.
2. Update $t_v = \min\{H'(b) : H'(b) > t_v\}$.
3. Morphologically reconstruct the collagen peak mask (i.e., a mask containing only one “1” pixel at the location of the collagen peak) into a $H' > t_v$ mask using the 4-neighborhood. If in the resulting mask both peaks belong to the same region, repeat steps 2–3.

We then set all bins in H' larger than t_v to t_v and use the inverse of such a histogram as the input for the watershed transformation; in the transform we fill holes in catchment basins.

Afterwards, we identify the basin encompassing the collagen peak (a binary mask) and one after the other:

1. Exclude 10% of its bins with lowest number of pixels (in H' histogram) located in the disk-shaped neighborhood of radius equals to the half of distance between the saddle point and the “other” peak, centered at the saddle point. The reason for this step is that it is the region where distributions corresponding to collagen and basophilic structures overlap, and thus the basin encompasses too many bins towards the “other” peak.
2. Close it with a $SE_D(1)$ structuring element, to fill the gaps between basins belonging to the collagen group (caused by dams).
3. Include bins to the left from the line $(0,0) - (C_{Col}^H, C_{Col}^E) - (C_{Col}^H, C_{RefMax}^E)$.
4. Exclude bins to the right from the diagonal $C^H(b) = C^E(b)$.
5. Exclude bins for which $C^H(b) > C_{Oth}^H$.
6. Exclude bins to the right of the line having the slope C_{Col}^E/C_{Col}^H and passing through the saddle point (with a margin one bin wide), assuming the collagen peak will not be excluded.
7. Grow all other basins by 2 bins and exclude those regions from the collagen’s “zone of influence”.

Pixels belonging to the remaining bins are treated as collagen. The binary mask consisting of pixels considered collagen will be further denoted as M_{Col} .

Finally, to coarsely reject regions of tissue which are not part of epidermis (unless the specimen was stained really poorly) in Section 5.1.4, we compute a binary mask M_{LowCH} of pixels with extremely low hematoxylin concentration by applying the following heuristic to each pixel P of hematoxylin concentration map:

$$M_{\text{LowCH}} = \begin{cases} 1, & \text{if } C^{\text{H}}(P) < \frac{1}{2} \min(C_{\text{NucMin}}^{\text{H}}, C_{\text{Col}}^{\text{H}}) \\ 0, & \text{otherwise} \end{cases}, \quad (5.7)$$

where $C_{\text{Col}}^{\text{H}}$ is hematoxylin concentration of the collagen peak.

5.1.4. Final refinement

In this stage we obtain the final epidermis segmentation, M_{Ep} , by rejecting false positive regions detected in the course of the stage 5.1.3 followed by performing the same set of operations as described in the step *Detecting epidermis* of the stage 5.1.2 (to retain only cohesive parts of the segmented epidermal region) and filling narrow gaps between individual epidermal regions.

The input to the aforementioned set of operations is the mask $M_{\text{SegPor}} \cap \overline{M_{\text{Red}}} \cap \overline{M_{\text{LowCH}}} \cap \overline{M_{\text{Col}}}$. As much of the stratum corneum layer is removed from the segmentation mask (it is included in the M_{Red} mask), the proximity region width is adjusted accordingly and set to $d = d_{\text{SCMax}} + r_{\text{EpidermMin}}$. If all regions in the resulting mask make less than p_{EpMin} of total lesion's area, it suggests that the collagen detection procedure failed (the M_{Col} mask contains too much false positives). In such case the mask is re-computed using the same set of operations and parameters as above, but this time for the mask $M_{\text{SegPor}} \cap \overline{M_{\text{Red}}} \cap \overline{M_{\text{LowCH}}}$ (i.e., regions detected as collagen are not removed). The result of this step will be denoted as M_{Ep0}

Lastly, we fill gaps between adjacent epidermal regions. Such gaps are often caused by slide preparation artifacts, such as mechanical tearing of the tissue (as discussed in section 2.2.4). This step is important for measuring epidermal morphometry, as one long section of epidermis is much more informative than several shorter ones (even if they have similar total length). To identify the gaps, we dilate each individual epidermal region with a disk-shaped structuring element of radius $2r_{\text{EpidermMin}}$ and mark overlapping regions in the M_{Gaps} mask. We then obtain the final epidermis segmentation, M_{Ep} according to the following formula:

$$M_{\text{Ep}} = M_{\text{Ep0}} \cup M_{\text{Gaps}} \cup (M_{\text{Ep0}} \oplus \text{SE}_{\text{D}}(2r_{\text{EpidermMin}}) \cap M_{\text{Gaps}} \oplus \text{SE}_{\text{D}}(2r_{\text{EpidermMin}})). \quad (5.8)$$

5.2. Results

5.2.1. Parameters selection

Our method has a few important parameters, including the minimal “radius” of the epidermal region, minimal solidity and maximal eccentricity of a clear cell region, and the minimal hematoxylin concentration in keratinocyte nuclei. To determine their values we selected 8 images of various conditions from our dataset as training images.

The threshold values in solidity- and eccentricity-based criteria used to identify valid clear cell candidate regions and actual clear cell regions (see the step of the Porosity analysis stage) were chosen based on the analysis of 200 examples of clear cell and non-clear cell void regions using a decision tree algorithm.

The minimal thickness of the epidermal plate, observed by a pathologist, was 20 μm (the plate was formed by four layers of flattened keratinocytes in granular layer). Similarly, the thinnest rete ridge was formed by two layers of average-sized keratinocytes. Therefore, the minimal “radius” of the epidermal region has been set to $r_{\text{EpidermMin}} = 10 \mu\text{m}$.

The minimum tissue density of epidermis after filling clear cells and desmosomes, $\rho_{\text{EpidermMin}}$, has been determined based on the epidermal region manually segmented by a pathologist. The value has been set to $\rho_{\text{EpidermMin}} = 0.95$ (an ideally processed epidermis would typically have a density of $\rho = 1$).

The minimal hematoxylin concentration observed by a pathologist in a typical keratinocyte nuclei was $C_{\text{NucMin}}^{\text{H}} = 0.15 C_{\text{RefMax}}^{\text{H}}$ (where $C_{\text{RefMax}}^{\text{H}}$ is a reference value of the maximum hematoxylin concentration). Threshold values for the detection of blood and stratum corneum were determined analogously. As stain concentrations are both determined automatically for each individual image and normalized, we assume that the observed values of those parameters will be valid for a wide range of images.

Even though values of the aforementioned parameters were determined by means of the visual examination of the training set, the range of their acceptable values (for each of those parameters) is narrow and strictly governed by morphological and biochemical properties of either tissue structures or individual cells.

5.2.2. Evaluation metrics

We assessed performance of the proposed algorithm using the three area-based performance metrics – sensitivity (\mathcal{A}_{SEN}), specificity (\mathcal{A}_{SPE}) and precision (\mathcal{A}_{PRE}) – defined as follows:

$$\mathcal{A}_{\text{SEN}} = \frac{|\text{GT} \cap \text{SEG}|}{|\text{GT}|}, \quad (5.9)$$

$$\mathcal{A}_{\text{SPE}} = \frac{|\overline{\text{GT}} \cap \overline{\text{SEG}}|}{|\text{GT}|}, \quad (5.10)$$

$$\mathcal{A}_{\text{PRE}} = \frac{|\text{GT} \cap \text{SEG}|}{|\text{SEG}|}, \quad (5.11)$$

where $|\cdot|$ is the cardinality operator, and GT and SEG denote the ground truth mask and the segmentation mask obtained by the automated technique, respectively. The ground truth mask for each WSI was prepared manually by an experienced pathologist with expertise in skin diseases.

5.2.3. Segmentation performance

We compared performance of the proposed method with the existing epidermis segmentation techniques, including the CET [111], GTSA [112], and THM [113].

The CET technique has several key parameters including low output thresholds for contrast enhancement, sizes of smoothing mean filter and morphological operations, and thresholds to eliminate noisy regions after thresholding. We set them to values proposed in the original paper [111]. The GTSA technique requires adjusting values of the following parameters: a window size of the template (T_{wsize}), a threshold for the shape criteria (T_1), a sampling resolution (T_{samres}), and a proportion of epidermis in a WSI (T_{pea}). In order to determine optimal parameters we evaluated performance of the GTSA technique on 10 images in the following feature space: $T_{\text{wsize}} \in \{3, 5, \dots, 9\}$, $T_1 \in \{2, 3, \dots, 9\}$, $T_{\text{samres}} \in \{0.005, 0.01, \dots, 0.1\}$, $T_{\text{pea}} = 0.01$ (the epidermis typically occupies at least 1.5% of a lesion, as described in Section 2.4.5, whereas a lesion typically occupies approximately 2/3 of the WSI crop area). Best results were obtained using the following values of parameters: $T_{\text{wsize}} = 3$, $T_1 = 2$, $T_{\text{samres}} = 0.005$. The key parameters of THM technique are thresholds for eliminating false positive regions and a threshold to determine the coarse segmentation quality. We set them according to the original work [113].

The presented method has been evaluated on 75 skin WSIs from the JUMC-SS dataset. The statistical results of quantitative evaluations using measures defined by equations 5.9–5.11 are shown in Table 5.1. Additionally, the number of images for which a given method failed (i.e., $|\text{SEG}| = 0$) has been provided. Note that the Table 5.1 summarizes mean and median values as well as inter-quartile range of the respective quantities computed only for the subset of images for which a given method did not fail. Figure 5.11 demonstrates intermediate results of the proposed segmentation pipeline.

Table 5.1. Performance evaluation of the epidermis segmentation on 60 images from the UJCM-SS dataset

Technique	#Failed	Mean value			Median value			IQR		
		\mathcal{A}_{SEN}	\mathcal{A}_{SPE}	\mathcal{A}_{PRE}	\mathcal{A}_{SEN}	\mathcal{A}_{SPE}	\mathcal{A}_{PRE}	\mathcal{A}_{SEN}	\mathcal{A}_{SPE}	\mathcal{A}_{PRE}
CET [111]	2	0.73	0.81	0.20	1.00	0.83	0.15	0.96	0.26	0.26
GTSA [112]	0	0.99	0.75	0.32	0.99	0.86	0.32	0.02	0.22	0.29
THM [113]	28	0.66	0.99	0.74	0.83	0.99	0.76	0.58	0.01	0.14
the proposed	0	0.97	0.97	0.69	0.98	0.98	0.75	0.04	0.02	0.17

5.3. Conclusions and discussion

The performance of the proposed technique is superior to the existing ones. Even though its mean sensitivity is comparable to that of the GTSA technique (0.97 vs. 0.99), it provides superior specificity and precision (0.97 vs. 0.75 and 0.69 vs. 0.32, respectively). Both CET and GTSA techniques have low precision, whereas the THM technique has an unacceptably high rate of failures (28 out of 75). On the other hand, only our method and the GTSA technique successfully yielded meaningful results in all 75 cases.

The poor performance of GTSA and THM techniques (i.e., low specificity and precision of the GTSA technique and high failure rate of the THM technique) results from the fact, that they are neglecting differences in staining protocols between laboratories – they are both based on assumptions that epidermis is uniformly stained and that in the red channel image there is a significant contrast between (dark) epidermis and (bright) collagen. However, the bulk of images in our dataset does not meet these assumptions: the contrast between epidermis and collagen is much lower, and there are intra-image variations in staining. As the result, Otsu’s thresholding produced two clusters of pixels (later used in the coarse segmentation step): the one consisting of darker pixels forming nuclei and stratum corneum, and the other one consisting of brighter pixels forming cytoplasm and collagen bundles. These darker regions were sparse and failed to satisfy area and shape criteria used in that segmentation step.

All the compared methods have relatively low precision, as numerous “dermal” pixels are incorrectly classified as epidermis in specimen with an abundant infiltration of cells (mainly lymphocytes and/or nevi cells) or with numerous skin appendages in dermis. Since cell nuclei appear dark purple, global thresholding (used in CET, GTSA and THM) incorrectly recognizes those low-intensity areas as part of epidermis. Additionally, in CET technique during the morphological processing step small false positive regions are often merged together into a larger clusters, which are not removed in the subsequent object classification step. Our technique fails to reject extremely dense connective tissue (like homogenized collagen) and cellular infiltrations close to lesion boundaries: both these tissue structures have hardly any pores, have stain concentrations similar to epidermis, and are either adjacent to epidermis or to lesion borders, therefore they are not rejected during any of the segmentation steps. In the THM algorithm the false positive area is limited by performing an additional segmentation step if the thickness of a given epidermal segment is too large, whereas the proposed technique achieves a similar goal by rejecting regions identified as epidermis, but which are confined within dermis.

Our method strongly relies on domain-specific knowledge of cell and tissue morphology and their biochemical properties. The porosity analysis stage is based exclusively on the analysis of void spaces; in particular in its morphological operations it utilizes only such well-defined quantities as radii of epidermal cells and their nuclei. The stain concentrations analysis stage is based on the analysis of normalized stain concentrations maps determined automatically for each individual image as well as it utilizes knowledge about the way certain tissue structures absorb both hematoxylin and eosin. Therefore, our technique is robust to differences in staining protocols, stain absorption within individual images, and

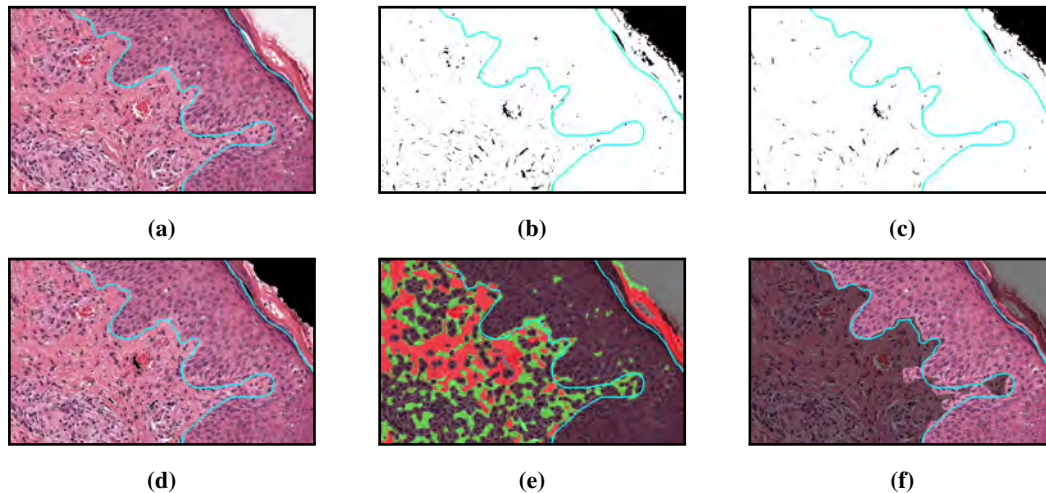


Figure 5.11. Intermediate results of the proposed epidermis segmentation pipeline for a skin lesion with scabies (Case 090, UBC dataset): (a) input image with ground truth marked blue, (b) specimen segmentation (slide foreground), (c) specimen segmentation after clear cells and desmosomes are filled, (d) coarse epidermis detection (rejected regions marked black), (e) stain concentrations analysis (red area – stratum corneum, blood or thick collagen; green area – collagen), (f) the final automatically determined epidermis segmentation (highlighted).

to illumination variability. Moreover, in the proposed method no assumptions about the shape of epidermis are being made – false positive regions are rejected only based on their position relative to lesion’s boundaries (i.e., if they are confined within dermis) – making it capable of segmenting epidermis in lesions with a wide range of inflammatory skin diseases, even when the morphology and structure of such lesions have broken down.

Although our technique yielded satisfactory segmentation result on 75 WSIs from our dataset, there are still some issues which need to be considered in order to apply the technique in routine clinical practice. Firstly, since dermal regions with abundant infiltration of cells with nuclei (in particular lymphocytes and nevi cell) are characterized by low porosity, and since the distribution and concentrations of hematoxylin and eosin stains in regions of infiltration are similar to those observed in epidermal region, the infiltrated regions are not rejected in any of the segmentation stages. Secondly, the presence of skin appendages in dermis, which is typical in both healthy and pathologically changed sections of skin, decreases precision rate of epidermis segmentation. The proposed method should also be verified on images of lesions containing combinations of pathological changes discussed in Section 2.5.

The proposed method for epidermis segmentation has a wide range of application. It could be used for detection of epidermal melanocytes, for diagnosis of inflammatory skin conditions, and to help find correlation between dermoscopic and histo(patho)logical features of physically damaged skin (for instance due to injuries [128]). We use it as an initial step in an algorithm measuring epidermal morphometry (described in Chapter 6).

5.4. Summary

In this chapter we presented an algorithm for epidermis segmentation in high-resolution WSIs of H&E-stained skin specimens. It performs an initial segmentation based on the analysis of shape and distribution of void spaces in a tissue, followed by the analysis of H&E stain concentrations in normalized stain concentrations maps obtained using color deconvolution. Finally, the segmented regions are verified against porosity and location constraints put on the epidermal region. We provided the segmentation results, and compared them both quantitatively and qualitatively with existing methods for epidermis segmentation.

The novelty of the proposed method is in its full automation, its robustness to color variations in staining, and its solid basis in skin histology and dermatopathology. Our method does not require any parameter selection from its users, as the relevant values are either computed from the slide image or predefined based on domain-specific knowledge. Since the applied approach is based on the analysis of images with normalized color, it is robust to inter-image and intra-image color variations due to non-uniform absorption of the stain, different handling procedure, and stains fading. The proposed method can be used, as a preliminary step, by higher-level algorithms used for the analysis on human skin, e.g., for the detection of epidermal melanocytes or for the measurement of epidermal morphometry. Therefore, it is a useful fundamental module for the development of CAD systems for dermatopathology.

6. Epidermis morphometry

In this chapter we present a novel method for automatic determination of the basic morphometric features of rete ridges in a segmented epidermis: their location, base width, length, and height. These information are then used to develop complex indexes describing the epidermal morphometry which in turn constitute input to the lesion classification model (as described in Chapter 7). The proposed method is the enhancement of our method described in [26]: it includes small (yet important) tweaks in all main steps of the algorithm so that it can be used not only with manually segmented epidermis masks, but also with the results of automatic epidermis segmentation methods (such as the one described in Chapter 5).

6.1. Algorithm for retes segmentation and for their morphometry measurement

The fundamental task in the presented method is to determine the base of each individual rete (Fig. 6.1a). The approximated location of retes along the epidermis main axis, henceforth referred to as “rete roots”, is identified by finding such projections of the skeleton of the epidermal region, which are branching off towards the underlying skin layers, or equivalently: not towards the epidermis “outer” edge (Fig. 6.3e, 6.3f). The “outer” edge of the epidermis is identified primarily by analyzing, for each point of the epidermis boundary, the distance to the slide background along the normal and in the opposite direction. The epidermis main axis is determined by finding such a path between two most distant endpoints in the epidermis skeleton which runs closest to the epidermis “outer” edge. Endpoints constituting rete bases are chosen by analyzing the curvature of the epidermis boundary to identify endpoint candidates, which are then matched with rete roots according to geometric criteria. Since typically a number of automatically segmented retes contain multiple individual tips whereas others are partially joined with their neighboring retes, a post-processing is necessary: the segmented retes must be merged together or split, respectively. After rete bases are identified, to increase the accuracy of locating rete roots, the epidermis main axis is retouched, so that it runs only through the epidermis base – and not through upper parts of retes (Fig. 6.3h). Finally, it is possible to measure the morphometry (width, length and height) of individual retes (Fig. 6.1b). The pipeline of our method is presented in Figure 6.2, whereas results of each of its steps are shown in Figure 6.3.

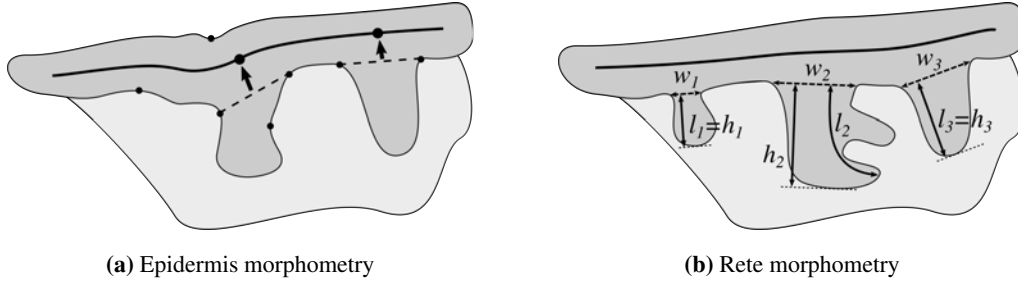


Figure 6.1. The morphometric features related to the epidermal region (dark gray): (a) the epidermis main axis is marked with the thick solid line, the bases of retes are marked with dashed lines, their corresponding roots as points on the epidermis main axis, and candidates for rete bases as points on the epidermis border; (b) base width (w), length (l), and height (h) of individual retes.

The foreground (tissue) mask M_{Fg} is obtained using the method described in Section 4.1. We also compute the M_{BgOut} mask representing the “outer” slide background by performing the morphological opening of the $\overline{M_{Fg}}$ mask with a disk-shaped structuring element of radius $8\ \mu\text{m}$ and rejecting all regions except the largest one. The epidermis mask M_{Ep} is obtained using the method described in Section 5.1, followed by filling holes smaller than $1000\ \mu\text{m}^2$ (which are most often pores in the stratum corneum).

6.1.1. Determining the epidermis “outer” edge

By the “outer” edge of the epidermis we denote the section of the epidermis boundary adjacent to stratum corneum. In order to identify the epidermis “outer” edge one needs to be able to distinguish between the “outer” direction (the same as the normal direction) and the “inner” one. Therefore, the initial step is to determine normals to the epidermis boundary.

To compute normals to the epidermis boundary we firstly get a sequence of border points P_1, \dots, P_n by tracing the border L_b in the clock-wise direction using Moore-Neighbor tracing algorithm modified by Jacob’s stopping criteria [129]. Next, to approximate the normal \vec{n}_i at the i th border point P_i we find the slope of the vector $\overrightarrow{P_{i-d}P_{i+d}}$ using the four-quadrant inverse tangent, and add $\pi/2$ to obtain a vector both orthogonal to $\overrightarrow{P_{i-d}P_{i+d}}$ and pointing “outside” the epidermis (Fig. 6.4). As the border is a closed shape, if $i-d < 0$ we take the point P_{n+i-d} and if $i+d > n$ we take the point P_{i+d-n} . Results are made more robust to minor local fluctuations in border’s curvature by setting the point distance to $d = \lceil 2\ \mu\text{m}/R_I \rceil$.

We identify the “outer” edge by casting two rays for each boundary point P_i : one in the same direction as its normal (\vec{r}_n) and another in the opposite direction (\vec{r}_o). For each ray we determine the distance along the ray to its first intersection with the slide background, d_{BgN} and d_{BgO} , respectively (Fig. 6.5a). If the WSI is cropped in such way that the tissue is not fully encompassed with the slide background, a ray will not cross the background – in such case we set the distance to be infinitely large. Due to the presence of numerous void spaces in dermis (between bundles of collagen fibers), the distance to the background is determined not using the “raw” mask, $M_{Bg} = \overline{M_{Fg}}$, but M_{BgOut} .

The point $P_i \in L_b$ is initially treated as part of the “outer” edge providing the following constraints are satisfied: (1) $d_{BgN} < d_{BgO}$, and (2) \vec{r}_n does not cross epidermis (Fig. 6.5a). We perform the above-

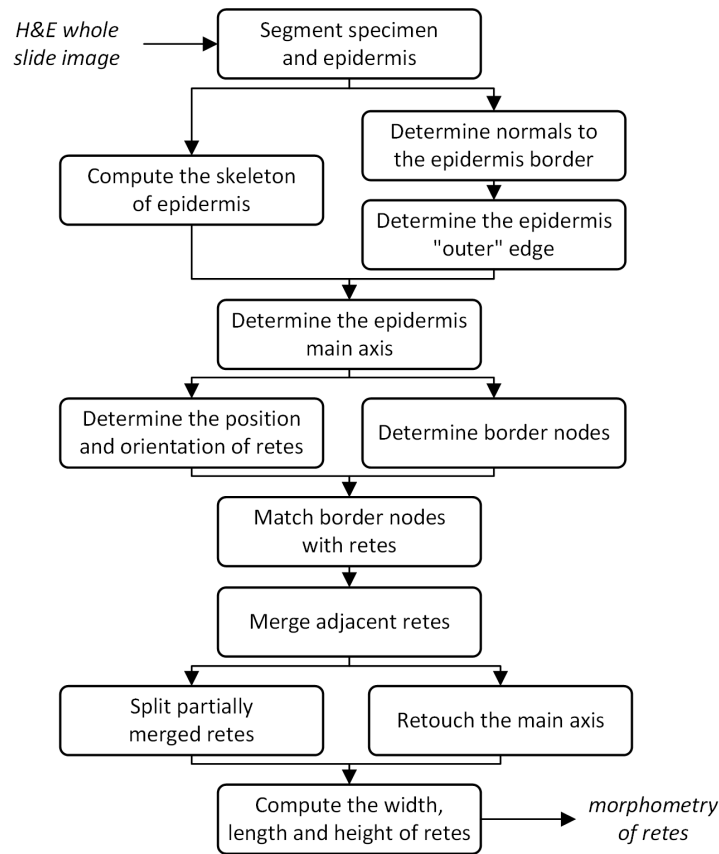


Figure 6.2. The proposed method pipeline

mentioned test for all boundary points and obtain a vector of logical values, $M_{\text{Out}0}$, which might be interpreted as a 1D binary mask.

As the initial constraints fail to correctly identify “pockets” in epidermis (Fig. 6.5b), we refine the $M_{\text{Out}0}$ mask by performing the following three steps (we will denote the results of the test “the \vec{r}_n vector crosses epidermis” as $M_{\text{Ep}X}$). Firstly, we identify all sequences of zeros in the $M_{\text{Ep}X}$ mask. For each such a sequence (spanning from the i th to the j th element), if both $(i - 1)$ th and $(j + 1)$ th element in the $M_{\text{Out}0}$ mask are ones, such a sequence is treated as a “pocket” constituting part of the valid “outer” edge (Fig. 6.5c). Then, we perform the following sequence of morphological closings and openings in order to both fill tiny gaps between parts of the valid “outer” edge and to get rid of artifacts:

$$M_{\text{Out}} = \left(\left(\left(M'_{\text{Out}0} \bullet \text{SE}_D(25\mu\text{m}/R_I) \right) \circ \text{SE}_D(250\mu\text{m}/R_I) \right) \bullet \text{SE}_D(650\mu\text{m}/R_I) \right) \circ \text{SE}_D(900\mu\text{m}/R_I), \quad (6.1)$$

where $\text{SE}_D(n)$ is a disk-shaped structuring element of radius n (the sizes of structuring elements were selected based on the observation of typical width of “pockets” between rete tips and between whole retes). Finally, we reject all but the largest segment (Fig. 6.5d).

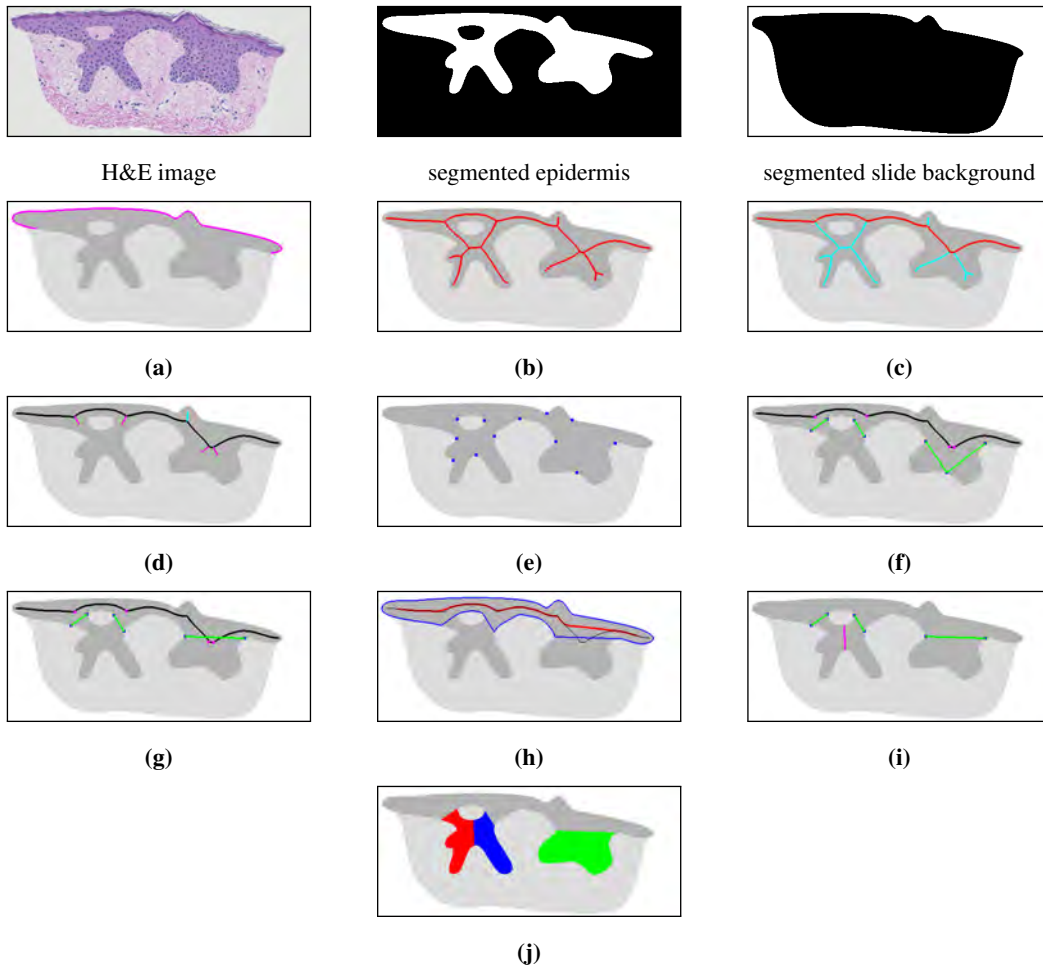


Figure 6.3. The steps of the proposed algorithm: (a) identify the “outer” edge of the epidermis, (b) compute the skeleton of the epidermal region, (c) identify the (initial) epidermis main axis (marked red), (d) find out the position and orientation of projections (“left side” and “right side” projections are marked cyan and magenta, respectively), (e) identify boundary nodes, (f) match boundary nodes with rete roots (rete bases and their roots are marked green and magenta, respectively), (g) merge adjacent rete roots, (h) recompute the epidermis main axis (the outline of the epidermis base is marked blue, the initial main and retouched main axis are marked black and red, respectively), (i) split partially merged retes (the split is marked magenta), and (j) compute the morphometry of each individual rete ridge (here marked with different colors for readability).

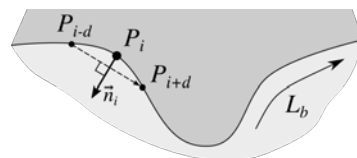


Figure 6.4. Approximating a normal to the epidermis boundary at the point P_i : the normal \vec{n}_i is orthogonal to the vector $\overrightarrow{P_{i-d}P_{i+d}}$.

6.1.2. Computing the epidermis skeleton

The skeleton of the epidermal region M_{Skel} is obtained in the following two steps. Firstly, we skeletonize the epidermal region (opened using a disk-shaped structuring element of radius $d_{\text{EpMin}}/2$ using an

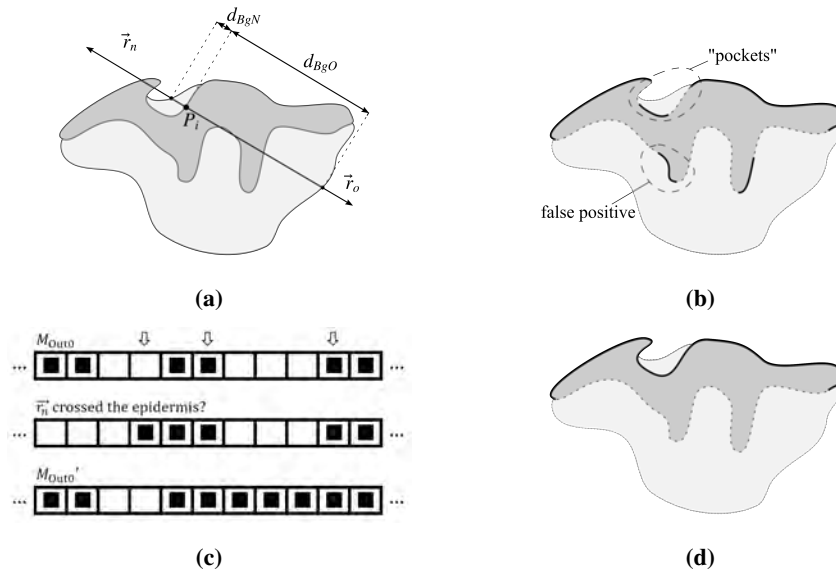


Figure 6.5. Determining the epidermis “outer” edge: (a) casting rays \vec{r}_n and \vec{r}_o for each point P_i and computing the distance to the background for each of those rays, (b) the initially identified “outer” edge mask M_{Out0} (thick solid line) may contain false positives, whereas some sections of the actual “outer” edge (e.g., “pockets”) may not be correctly identified, (c) considering all such sequences of 0s in the M_{Out0} mask that both for all their elements the \vec{r}_n vector crosses the epidermis and which are flanked with 1s in the M_{Out0} mask as parts of the valid “outer” edge yields the M_{Out0}' in which most issues with “pockets” are fixed, and (d) performing the sequence of morphological closings and openings on the M_{Out0}' mask fixes the remaining issues with “pockets” and false positives – the final version of the epidermis “outer” edge, the M_{Out} mask, is marked with a thick solid line.

augmented fast marching method proposed by Telea and van Wijk [130] with pruning threshold t . The opening operation is performed to limit the influence of boundary artifacts related to imperfect segmentation of the stratum corneum). Then, we additionally apply the medial axis transform to the resulting mask in order to remove excessive pixels near junctions.

The pruning threshold t has a precise geometrical meaning: all skeleton branches caused by boundary details shorter than t pixels are pruned [130]. In practice, even an expert pathologist is able to reasonably conclude whether a given boundary detail could be a rete only if that detail has the length of at least two average-sized keratinocytes. As the minimum thickness of the epidermal plate found in human body varies between 45–65 μm [76, 77] and as the diameter of an average keratinocyte is 12–16 μm [74], the distance from the epidermis main axis to the tip of a boundary detail should be at least about 50 μm . Nevertheless, to avoid missing potential retes located at abnormally thin epidermal sections (e.g., in melanoma) we adopted the pruning threshold value $t = 45 \mu\text{m}/R_I$ (Fig. 6.6).

6.1.3. Determining the epidermis main axis

The epidermis main axis π_m is a centerline of the epidermis base, whereas the epidermis base is the epidermal region with retes cut off (Fig. 6.1a). The determination of the main axis allows for the

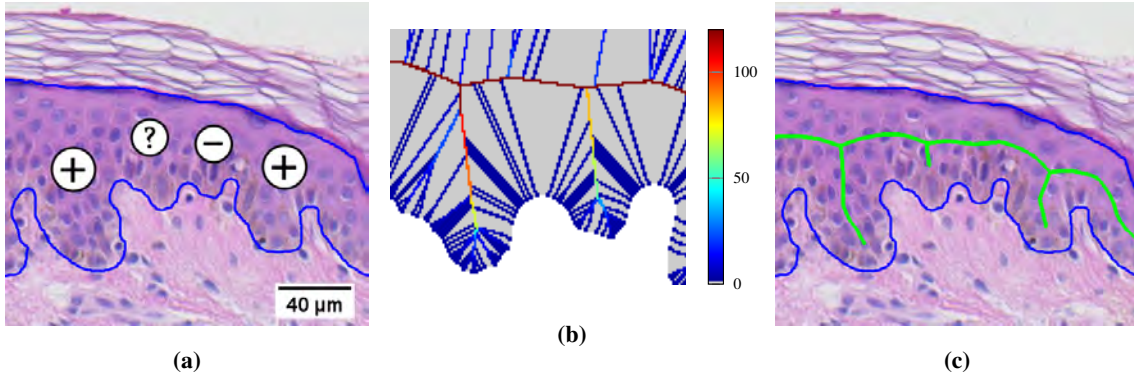


Figure 6.6. A section of epidermis: (a) two boundary details are definitely retes (“+”), one could be a rete (“?”), and one is surely not a rete (“-”); (b) the U difference field for the two central boundary details (see [130] for details); (c) the skeleton obtained by thresholding the U difference field with $t = 100$.

identification of pixels being candidates for rete roots (since all roots of retes are located on that axis) and for distinguishing between epidermal projections branching off towards the deeper skin layers and towards the stratum corneum.

The procedure for determining the epidermis main axis is as follows:

1. We construct graphs G and G_f spanned on the skeleton of the “raw” epidermal region and the skeleton of the epidermal region with void spaces filled, respectively (skeletonization is performed using the procedure described in Section 6.1.2). The leaves of these graphs represent skeleton endpoints. The location (in the image coordinate system) of leaves in G corresponds to the location of leaves in G_f , as the filling operation only removes skeleton branch points.
2. The naive approach to the problem of determining the epidermis main axis – finding the longest path in G – has two problems: (1) if epidermis contains holes (e.g., between partially merged retes) the longest path in G may not necessarily run along the epidermis “outer” edge (Fig. 6.7a), and (2) simply selecting the longest path in G yields incorrect results for cases where there is a long rete near any end of the epidermis base (Fig. 6.7b). Therefore, we identify the two endpoints of the main axis by selecting the best path in the graph G_f . We rank paths in G_f according to the score which prefers longer paths but at the same time penalizes paths with endpoints located deep inside the epidermis (i.e., close to the dermis):

$$\text{Score}(\pi) = \sum_{e_i \in \pi} w(e_i) - (D_{\text{Out}}(v_0) + D_{\text{Out}}(v_n)) \quad (6.2)$$

where $\pi = (e_1, \dots, e_p)$ is a path in the graph G_f , $w(e)$ is the weight of the edge e (i.e., the number of pixels forming this edge), $D_{\text{Out}}(v)$ is the geodesic distance of v from the epidermis “outer” edge, and v_0 and v_n are endpoints of π . We compute $\text{Score}(\pi)$ for all such paths in G_f that v_0 and v_n are the leaves of G_f , and choose endpoints of the path with the highest score as the main axis endpoints v_1^e and v_2^e (Fig. 6.7c).

3. We construct a weighted graph G_w , which has the same nodes and edges as G , but every edge e in G_w is weighted according to both the mean depth of pixels P_i forming this edge (to prefer graph edges closer to the epidermis “outer” edge) and the presence of narrow passages on the edge (such

passages are typically artifacts due to improperly segmented stratum corneum):

$$\begin{aligned}
 w_d(e) &= \sum_{P_i \in e} D_{\text{Out}}(P_i) \\
 w_1(e) &= \begin{cases} \max(w_d(f) : f \in E(G)) & \text{if } \exists P_i \in e : D_{\text{Bg}}(P_i) < d_{\text{EpMin}}/R_I \\ 0 & \text{otherwise} \end{cases} \quad (6.3) \\
 w(e) &= w_d(e) + w_1(e)
 \end{aligned}$$

where $D_{\text{Bg}}(P_i)$ is an Euclidean distance transform of the M_{Bg} mask at the P_i pixel (i.e., the distance from the P_i pixel to the nearest background pixel). To ensure that the chosen route is running through the epidermis base, we determine the route of the main axis by searching for the longest path between v_1^e and v_2^e in G_w using Dijkstra's algorithm [127].

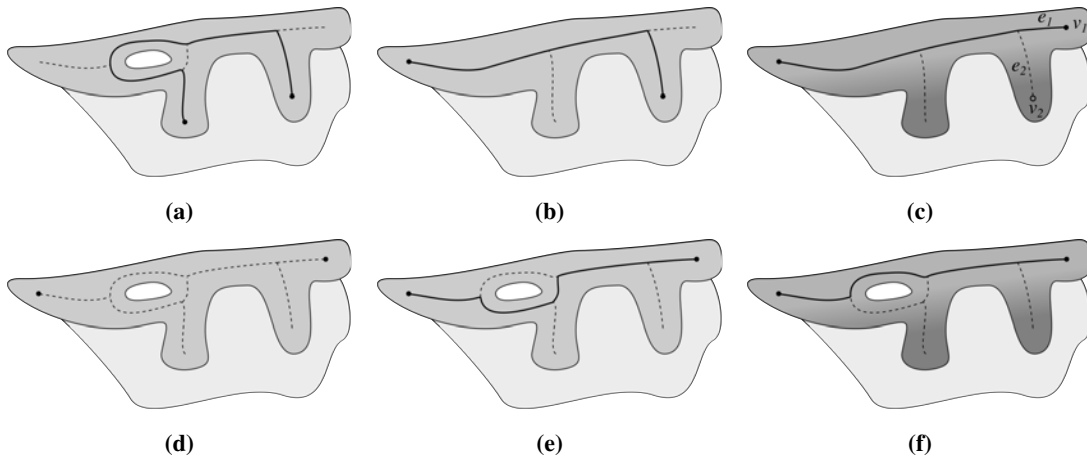


Figure 6.7. Determining the epidermis main axis – issues and their solutions (in each diagram the computed main axis and its endpoints are marked with a thick solid line and large dots, respectively): (a) due to holes in epidermis the longest path in G does not run along the epidermis outer edge, (b) although the problem with holes can be solved by identifying main axis endpoints based on the analysis of the skeleton of the filled epidermal region, this tweak does not solve the problem of long retes near ends of the epidermis base causing the longest path in G to omit the last segment of the true main axis, (c) the problem with long retes near ends of the epidermis base can be solved by ranking endpoints based on their depth (i.e., by preferring endpoints located closer to the epidermis “outer” edge), (d) endpoints determined using G_f are also valid for G , (e) simply taking the longest path in G between endpoints is not enough to obtain the correct main axis route, (f) finding the shortest path in G_w between endpoints yields the desired result.

6.1.4. Retes segmentation

The aim of this step is to determine position on the main axis and orientation of retes, as well as to identify endpoints of each rete base.

Determining the position and orientation of retes

The process of determining the location and the orientation of retes can be summarized in the following three steps, performed for each projection branching off from the epidermis main axis: (1) compute its

location and orientation, then (2) determine the side of the main axis it branches off from (either “left” or “right”), and finally (3) treat it as a rete only if it branches off from that side of the main axis which is closer to the epidermis “outer” edge. The detailed procedure is described below.

Firstly, to determine the location and the orientation of projections, for each such edge $e_i = (v_q, v_r)$ that $v_q \in \pi_m$ we choose v_q as the projection root. Then, to approximate the orientation φ_p of that projection, we use the following heuristic: we choose the pixel $v_p \in e_i$ located $L_p = \min(|e_i|, l_{\max})$ pixels away from v_q as the end of the projection “stump” and compute the four-quadrant inverse tangent of the $\overrightarrow{v_q v_p}$ vector (Fig. 6.8). We set a small value of $l_{\max} = \lceil 5 \mu\text{m}/R_I \rceil$ to be able to reasonably approximate the rete orientation even if the rete is significantly twisted closer to its tip. To get the “stumps”, we use a helper M_{Stumps} mask computed as follows:

$$M_{\text{Stumps}} = M_{\text{Axis}} \xrightarrow{8} (M_{\text{Axis}} \oplus \text{SE}_D(l_{\max})) \cap M_{\text{Skel}}$$

where M_{Axis} is the mask of the epidermis main axis.

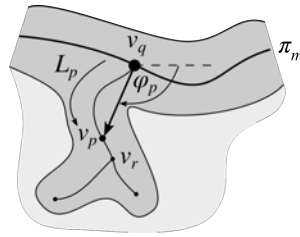


Figure 6.8. Determining the orientation φ_p of a rete based on the location of its root v_q and its corresponding end of a rete “stump” v_p .

Then, to determine the side of the main axis a given projection branches off from (either “left” or “right” – this labeling is arbitrary), we dilate the main axis using a disk-shaped structuring element of radius l_{\max} , split the resulting region in half using the main axis and two rays (the one cast from the first pixel of the axis in the direction opposite to the direction of the first axis segment, and the other cast from the last pixel in the same direction as the last segment), and check with which half the end of this rete “stumps” overlaps (Fig. 6.9).

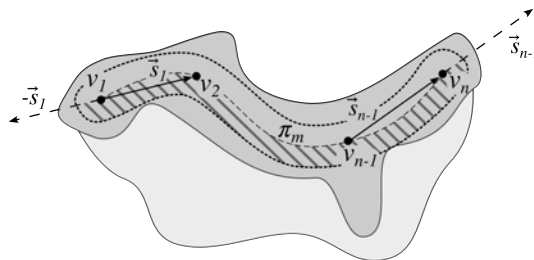


Figure 6.9. Determining side of the main axis (either “left” or “right”) a given projection branches off from. The dilated main axis region is split in halves by the main axis π_m combined with two rays: $-\vec{s}_1$ (cast from v_1) and \vec{s}_{n-1} (cast from v_n). If the projection “stump” is located in the striped half, it is considered a “right” projection, otherwise it is considered a “left” projections.

Finally, we treat as retes only those projections which branch off from the “outer” side of the main axis. In order to find out which side of the main axis (“left” or “right”) is the “outer” one, for each projection on each side we cast a ray in the same direction as the orientation of that projection and determine the

distance along that ray to its first intersection with the slide background. The side for which its projections has larger mean distance to the slide background is considered the “outer” one (and, consequently, these projections are considered retes).

Determining boundary nodes

The preliminary step towards the identification of endpoints of rete bases is the determination of candidate points on the epidermis boundary (a boundary point is labeled as a candidate if it is located in the center of a boundary protrusion). It is not enough to simply choose two points on the boundary closest to a given rete root as for wide, long retes the main axis traced using the skeleton is considerably deformed – its root is located in the center of the rete – and therefore we would choose points in the middle of rete’s length instead of points close to its base (Fig. 6.10).

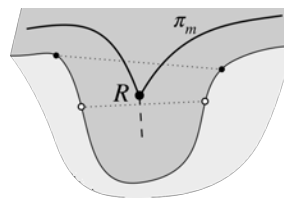


Figure 6.10. To identify endpoints of rete bases we firstly determine candidate points, located in the middle of boundary protrusions (black dots). Simply selecting two points (white dots) on the boundary closest to the rete root R gives wrong results for wide and long retes as in such case the main axis π_m traced using the skeleton is considerably deformed (the junction point R is shifted towards the center of a rete).

Firstly, for each boundary (both external and internal), we compute relative changes in the approximated boundary slope between consecutive boundary points (using the procedure described in Section 6.1.1) and represent it as a function of boundary run length (Fig. 6.11a). We assume that all boundaries are traced in the clock-wise direction and that boundaries consisting of less than 5 px are actually artifacts. The relative orientation of the vector v_1 with respect to the vector v_2 is given by:

$$[-\pi, \pi] \ni \Delta\varphi(v_1, v_2) = \begin{cases} \varphi_1 - \varphi_2 & \text{if } \varphi_1 - \varphi_2 \in [-\pi, \pi] \\ \varphi_1 - \varphi_2 + 2\pi & \text{if } \varphi_1 - \varphi_2 < -\pi \\ \varphi_1 - \varphi_2 - 2\pi & \text{if } \varphi_1 - \varphi_2 > \pi \end{cases} \quad (6.4)$$

where φ_1 and φ_2 are the absolute orientations of vectors v_1 and v_2 , respectively.

Then, tiny fluctuations in the function graph are suppressed by applying the simple moving average filter with window size $w_s = 15$ and the general trend is obtained by performing low-pass filtering using a minimum-order low-pass finite impulse response (FIR) filter with normalized passband frequency of 0.1 rad/px, stopband frequency of 0.15 rad/px, passband ripple of 0.01 dB, and stopband attenuation of 65 dB, designed using a Kaiser window method (for boundaries consisting of less than 100 pixels we use stopband attenuation of 50 dB). Since in both cases of filtering the signal corresponds to a value measured along the boundary (which is a closed curve), we wrap the signal to avoid filtering artifacts on its ends. The results of the compound filtering are shown in Figure 6.11b.

Finally, we identify the location of such local maxima (peaks) in the filtered (and wrapped) signal that their height h and prominence p are large enough, as higher peaks correspond to sharper turns of the boundary and more prominent peaks indicate a more constant trend (Fig. 6.11c). We chose $h > 0.025$ and $p > 0.0015$, respectively.

Since no “golden standard” for determining the precise location boundary nodes exists (it is an extremely subjective task), we selected the values of filter parameters and peak thresholds empirically.

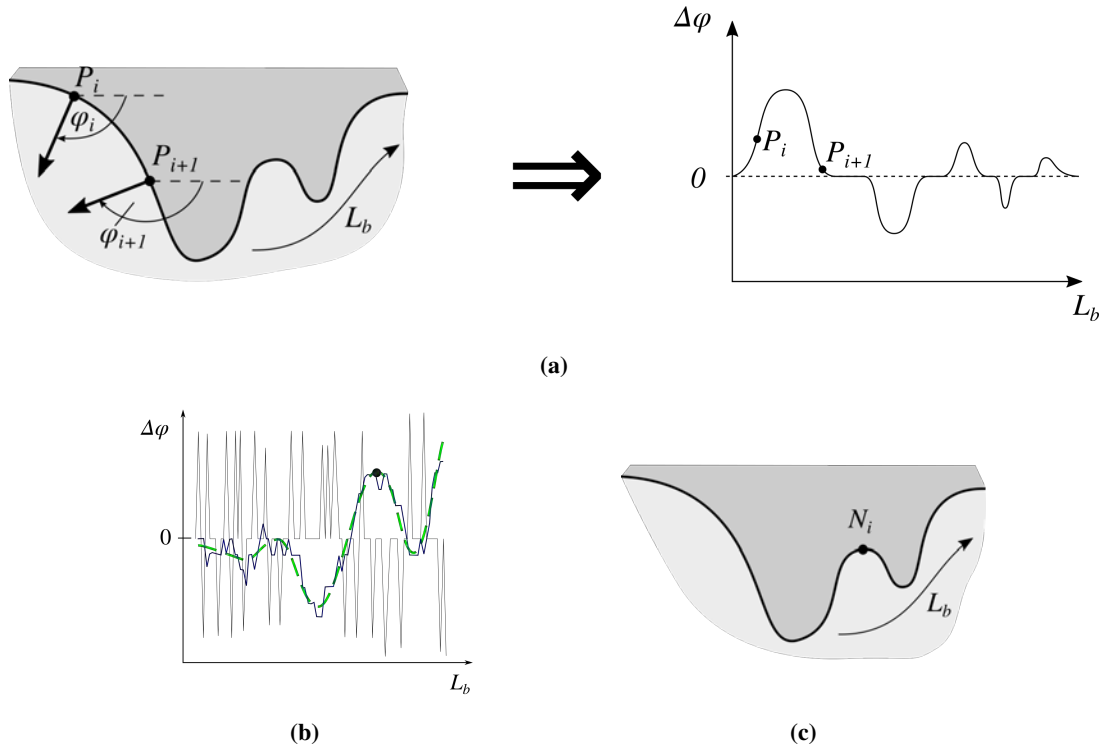


Figure 6.11. Identifying boundary nodes: (a) relative changes in the approximated boundary slope between consecutive boundary points are represented as a function of boundary run length $\Delta\varphi(L_b)$, (b) the $\Delta\varphi(L_b)$ function (thin solid line) is firstly averaged using the simple moving average filtering (thick solid line) and then a low-pass filter is applied (dashed line), (c) the identified node is located in the center of a boundary protrusion.

Matching boundary nodes with retes

For each rete root R we need to determine the two endpoints of its base. A candidate for each of such endpoints, N_i , must be located near the rete root (we set $\|RN_i\| < 130\mu\text{m}/R_I$ as a reasonable limit) and must be located on the epidermis “inner” edge. Next, candidate nodes are labeled as either “left” or “right” based on the following criterion: if $\Delta\varphi(\vec{n}_i, \vec{p}) > 0$ then the node is labeled as “left” candidate, otherwise it is treated as a “right” candidate (\vec{p} is a rete vector and $\vec{n}_i = \overrightarrow{RN}_i$). Within each category of candidates we firstly try to choose the node closest to R for which $|\Delta\varphi(\vec{n}_i, \vec{p})| \in (5^\circ, 150^\circ)$. If there are no candidates satisfying this criterion, we relax it and choose from all the nodes in the group (disregarding their $|\Delta\varphi(\vec{n}_i, \vec{p})|$ value).

Sporadically, the lack of concavities in epidermis boundary near a rete root may result in one of the following scenarios happening for one side of a rete: (1) a node too distant from the rete root is selected,

(2) a node on the wrong side of the main axis is selected, or (3) no node is selected. Figure 6.12 presents each of these scenarios.

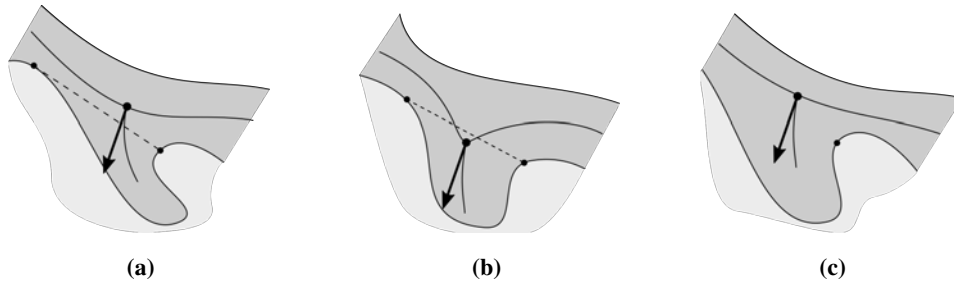


Figure 6.12. Problematic scenarios when identifying endpoints of a rete base: (a) a node too distant from the rete root is chosen, (b) a node on the wrong side of the main axis is chosen, and (c) no node is chosen.

Should any of the above-mentioned cases occur, we select the boundary point located on the ray cast from the rete root in the direction obtained by mirroring the direction from rete root to the “proper” node (i.e., the one located closer to the rete root if two nodes were initially found, otherwise the one found) with respect to the rete vector (Fig. 6.13).

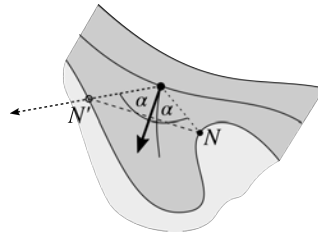


Figure 6.13. The “proper” node N is mirrored with respect to the rete vector to obtain a new node N' .

Merging adjacent retes

If at least one of the following conditions is met:

- the distance between the roots of two retes is less than $20\ \mu\text{m}$,
- two retes have exactly the same base endpoints, or
- one rete is fully enclosed by another rete

we treat them as a group and merge into a single, “joint” rete (a group may consist of more than two retes).

To determine the location and the orientation of such a joint rete, we take the average of the orientations of all retes in the group and the point R' on the main axis closest to the center of mass of roots R_{avg} of all retes in the group, respectively. To select base endpoints of the joint rete, within each group we firstly identify two most distant roots R_a and R_b . Then, for each node N_i constituting an endpoint of the base of a rete belonging to the group we calculate the relative orientation of the vector $\vec{n}_i = \overrightarrow{ON}_i$ (we choose the point O in such a way that $\|OR_a\| = \|OR_b\|$) with respect to the mean vector of rete orientations \vec{p}_{avg} according to (6.4), and finally we select one node with the largest and one node with the smallest relative orientation $\alpha = \Delta\varphi(\vec{n}_i, \vec{p}_{\text{avg}})$. Figure 6.14 shows the procedure of merging adjacent retes.

We repeat the above-mentioned merging steps until there are no more groups left.

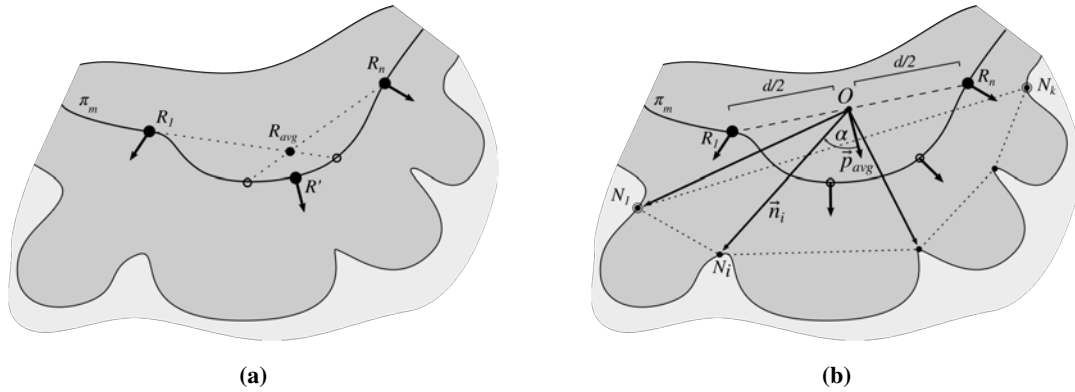


Figure 6.14. Merging the adjacent retes: (a) the root of the new (joint) rete R' is the point on the main axis located closest to the center of mass of joint roots R_{avg} , (b) the two most distant retes roots are R_1 and R_n and the base nodes of the new (joint) rete are N_1 and N_k .

Splitting partially merged retes

In case of lesions such as junctional nevus or melanoma adjacent retes are frequently partly joined together and thus must be (at least roughly) delimited in order to measure the height of each individual rete (Fig. 6.15).

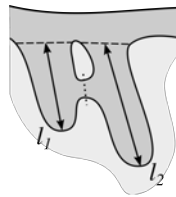


Figure 6.15. Joined retes must be split (the dotted line) before measuring their morphometry, otherwise they both would have the same length (in this case equals to l_2).

The process of delimiting retes begins with tracing boundaries of each hole in epidermis (i.e., a hole which is fully enclosed within the epidermis in an 8-connected neighborhood) and then repeating the procedure described below row by row (i.e., in the i th pass we consider only holes forming the i th row) until all boundaries are processed (Fig. 6.16).

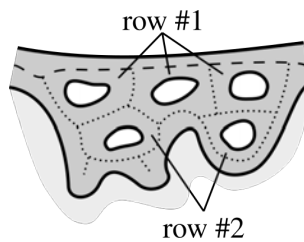


Figure 6.16. Rows of holes: the 1st row is formed by such cycles around holes that at least one edge in each cycle belongs to the main axis (a dashed line); the 2nd row is formed by cycles adjacent to the first row; the n th row is formed by those cycles adjacent to the $(n - 1)$ th row which are not part of any of 1st, 2nd, ..., $(n - 1)$ th row.

When processing a row, holes are analyzed one by one – in the j th iteration we begin with the identification of nodes in the M_{Skel} mask forming the cycle around the j th hole (we will denote these nodes as V_j): we reconstruct a mask containing the hole border into the $(M_{Ep} \setminus M_{Skel})$ mask, dilate the result with

a square-shaped structuring element of size 3×3 (so that it includes skeleton nodes), and perform the logical AND operation with skeleton nodes.

The splitting algorithm is as follows:

- In the first pass (i.e., $i = 1$) all holes for which their boundaries do not overlap with the main axis are skipped. For each of the rest of holes the split is performed in the following steps (Fig. 6.17):

1. Compute the “hole proximity” mask of a hole:

$$M_{\text{HoleProx}} = (M_{\text{Hole}} \oplus \text{SE}_{\text{Sq}}(3)) \stackrel{4}{\rightarrow} (M_{\text{Ep}} \cap \overline{M_{\text{Skel}}}) \oplus \text{SE}_{\text{Sq}}(3),$$

where M_{Hole} is the mask of a hole, and $\text{SE}_{\text{Sq}}(n)$ is a square-shaped structuring element of edge length n . The holes in M_{HoleProx} are filled afterwards.

2. Identify the two edges flanking a given hole (constituting “trunks” of two retes) by finding the outermost edges in the longest path in a graph spanned on the skeleton given by the $M_{\text{Stumps}} \cap M_{\text{HoleProx}}$ mask. Take the mean of their orientations as the orientation of the splitting ray.
3. Cast the splitting ray from this point on the hole boundary which is most distant (in Euclidean metric) from the epidermis “outer” edge – the ray splits the epidermis until it reaches the background region.

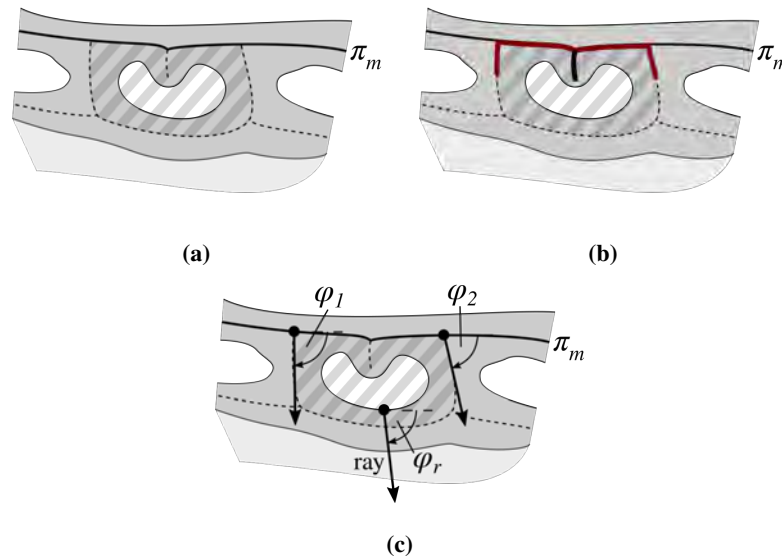


Figure 6.17. The splitting procedure for holes in the 1st row: (a) compute the “hole proximity” mask of a hole (it includes the adjacent sections of the epidermis skeleton), (b) identify the longest path (marked red) in a graph spanned on the skeleton given by the $M_{\text{Stumps}} \cap M_{\text{HoleProx}}$ mask (it will then allow for the determination of the two retes flanking a given hole), and (c) cast the splitting ray from this point on the hole boundary which is most distant (in Euclidean metric) from the epidermis “outer” edge (the orientation of the splitting ray equals the mean of orientations of retes “flanking” the hole).

- In each of the subsequent passes (i.e., $i > 1$), for each hole in a row:

1. Compute the “hole proximity” mask M_{HoleProx} as in the 1st pass.

2. Compute a helper mask $M_G = M_{\text{HoleProx}} \cap (N > 0) \cap M_{\text{Skel}}$. If $|M_G| = 0$, the hole will not be split at all; if the M_G mask contains more than two connected components, it means that another hole should be processed before the current one.
3. Find the longest path π_L in a graph G spanned on the M_G mask.
4. If the π_L path contains a cycle, treat it as a loop (Fig. 6.18):
 - (a) Compute the geodesic distance transform of the M_{Ep} mask with seeds located at the M_{Axis} mask, it will be denoted as $\text{Dist}_{\text{Axis}}$.
 - (b) Choose the “deepest” point in the hole P_{Deep} :

$$P_{\text{Deep}} = \arg \max \{ \text{Dist}_{\text{Axis}}(P) : P \in M_{\text{Hole}} \}.$$

- (c) Choose the point P_{Far} according to the following criterion:

$$X = \{ P \in I : \overline{P_{\text{Deep}}P} \cap M_{\text{Hole}} = \emptyset \}$$

$$P_{\text{Far}} = \arg \max \{ \text{Dist}_{\text{Axis}}(P) : P \in X \wedge \|\overline{P_{\text{Deep}}P}\| \leq 10\text{px} \}$$

That is, choose such a point that the line $\overline{P_{\text{Deep}}P_{\text{Far}}}$ does not cross the hole region.

- (d) Cast the splitting ray from the P_{Deep} point in the direction of the P_{Far} point – the ray splits the epidermis until it reaches the background region.
- (e) Continue with processing the next hole.

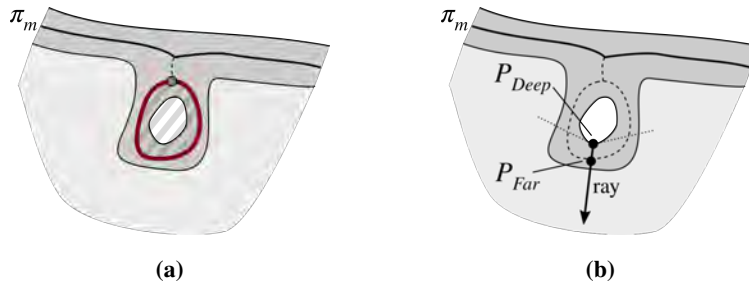


Figure 6.18. The splitting procedure for “looped” holes: (a) the π_L path contains a cycle, thus the hole is treated as “looped”, and (b) cast the splitting ray from the P_{Deep} point in the direction of the P_{Far} point (the dotted lines from the P_{Deep} point mark the “upper” boundary of the X pixel set).

Otherwise (if the π_L path was not treated as a loop):

- (a) Identify endpoint candidates – vertices in the graph G lying on these edges of G which do not constitute boundaries of the already-processed regions (i.e., candidates must belong to the $M_{\text{HoleProx}} \cap (N = 0)$ mask), excluding endpoints of edges consisting only of two pixels (i.e., spurs).
- (b) If no valid endpoint candidates were found, treat the hole similarly to a loop (Fig. 6.19):
 - i. Choose the hole point P_{Deep} most distant from the hole base, $M_{\text{HoleProx}} \cap (N > 0)$ (in the Euclidean distance).
 - ii. Take the orientation of the vector from the centroid of the M_{Hole} region to the P_{Deep} point as the orientation of the splitting ray and cast it from the P_{Deep} point.
 - iii. Continue with processing the next hole.

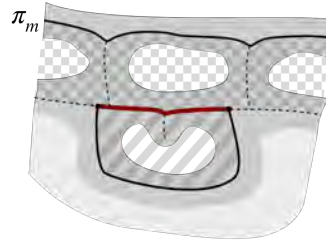


Figure 6.19. The “quasi-looped” holes in further rows: the only possible edge (marked with solid black lines) contains both endpoints of the π_L path (marked red), there are no endpoint candidates; the already-processed regions ($N > 0$) are marked with a checker pattern.

Otherwise (Fig. 6.20):

- i. Finds (two) edges containing both the one endpoint of the π_L path and an endpoint candidate.
- ii. If no such edges were found, treat the hole as if no valid endpoint candidates were found (i.e., follow substeps described in step 4b); otherwise take the mean orientation of edges as the orientation of the splitting ray.
- iii. Cast the splitting ray from the point most distant from the epidermis “outer” edge (in the Euclidean metric).

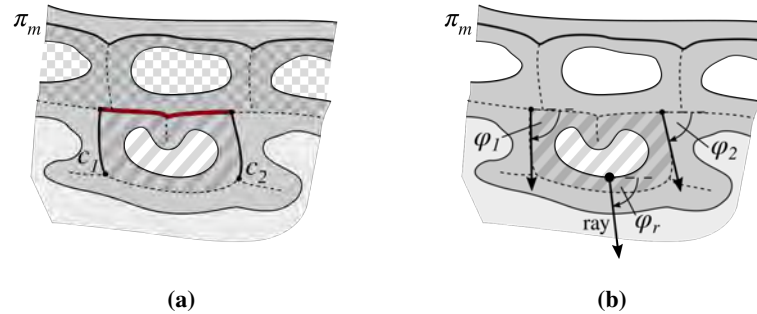


Figure 6.20. The splitting procedure for “regular” holes in further rows: (a) identify the two edges (marked with solid black lines) containing both the one endpoint of the π_L path (marked red) and an endpoint candidate (here there are two endpoint candidates, c_1 and c_2); the already-processed regions ($N > 0$) are marked with a checker pattern, and (b) cast the splitting ray from this point on the hole boundary which is most distant (in Euclidean metric) from the epidermis “outer” edge (the orientation of the splitting ray equals the mean of orientations of edges “flanking” the hole).

- If there was not a single change in the given step (i.e., not a single hole was actually processed), facilitate processing of potential H-sections of the skeleton by marking the node closest to regions in the M_{Axis} mask of the “deeper” hole, denoted as v_{Shallow} , as processed once, i.e., by setting $N(v_{\text{Shallow}}) = 1$ (the geodesic distance transform of the M_{Skel} mask with seeds located at the M_{Axis} mask is used to determine the v_{Shallow} point). A sketch of a H-section is shown in Figure 6.21.
- Regardless the current step number, after processing each hole the number of times a given image region has been processed is updated: $N(M_{\text{HoleProx}} \oplus \text{SE}_{\text{Sq}}(3)) = N(M_{\text{HoleProx}} \oplus \text{SE}_{\text{Sq}}(3)) + 1$.
- The remaining (unprocessed) holes are treated as loops (i.e., as described in step 4).

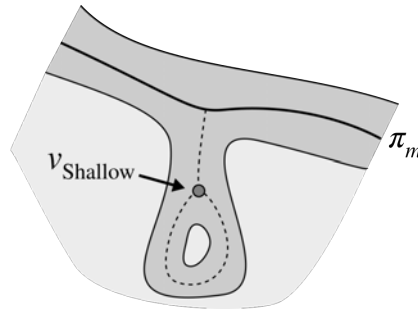


Figure 6.21. An example of a H-section of the skeleton. The v_{Shallow} node should be marked as processed once (by setting $N(v_{\text{Shallow}}) = 1$) in order to be able to process the hole at the tip of a rete.

The splitting procedure performed on holes from the 1st and the 2nd row is shown in Figure 6.22. The result of this procedure (performed on all rows) is a binary mask M_{Split} representing a superposition of rasterized images of all splitting rays (it is later used when computing base width and height of individual retes).

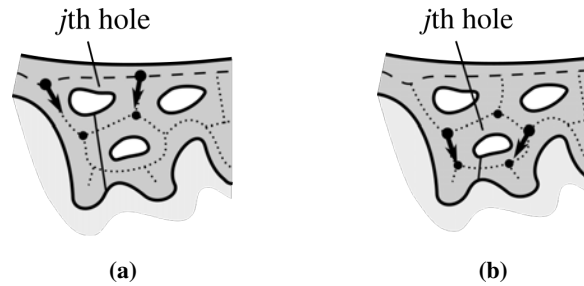


Figure 6.22. Delimiting partially merged retes: (a) a hole in the 1st row, (b) a hole in the 2nd row. Nodes V_j are marked with circles, the orientation of edges flanking the hole is marked with an arrow, and the splitting ray is marked with a solid line.

6.1.5. Retouching the main axis

As mentioned in Section 6.1.3, the initially determined route of the epidermis main axis may run through upper parts of broad retes, which negatively affects the algorithms for finding rete roots and rete bases. Therefore, after rete bases are determined, we recompute the main axis so that it runs only through the epidermis base.

To determine the epidermis base mask M_{EpBase} , we render the rete bases using Bresenham's line algorithm [131] into an (initially empty) mask M_{Bases} and then reconstruct the original main axis mask into the $M_{\text{Ep}} \setminus M_{\text{Bases}}$ mask (i.e., into the epidermis mask with rete bases removed). To obtain the retouched routing of the main axis, we repeat the procedure of identifying the "shallowest" path described in Section 6.1.3, this time on the skeleton of M_{EpBase} and with the same endpoints as in the original main axis. The skeleton is computed using the procedure described in Section 6.1.2, this time with the threshold t twice as large.

When re-computing the main axis the choice of the pruning threshold value t is more arbitrary, since in terms of the skeleton structure it only affects the length of sections of the main axis cut off from each

end of the axis – a larger value of t results in longer section to be cut off. However, it significantly affects the computational complexity of the main axis routing algorithm, as a larger value of t results in a graph with less nodes and edges to process.

6.1.6. Computing the morphometry of retes

In this step we measure the following morphometric parameters of each individual rete: base width, length and height.

To extract a mask of an individual rete, we firstly compute a mask $M_{\text{Proj}} = M_{\text{Ep}} \setminus (M_{\text{Base}} \cup M_{\text{Split}})$ containing all rete regions, and then reconstructing the edge of that rete into M_{Proj} . The base width of a rete is simply the length of its base (i.e., the Euclidean distance between base endpoints). The length of a rete is determined by computing the geodesic distance transform of the projection's region with seeds located at the projection's base and taking the maximum value found in the transformed image (Fig. 6.23a). To determine the height of a rete, for each boundary pixel P we measure its Euclidean distance from the line $k : Ax + By + C = 0$ containing the rete base and take the maximum value across all boundary points (Fig. 6.23b). Assuming $N_1 = (x_1, y_1)$ and $N_2 = (x_2, y_2)$ are endpoints of the rete base, the coefficients of the line k are given by the following equations:

$$A = \begin{cases} 1 & \text{if } x_1 = x_2 \\ \frac{y_1 - x_2}{x_1 - x_2} & \text{otherwise} \end{cases} \quad B = \begin{cases} 0 & \text{if } x_1 = x_2 \\ -1 & \text{otherwise} \end{cases} \quad C = \begin{cases} -x & \text{if } x_1 = x_2 \\ y_1 - \frac{y_1 - x_2}{x_1 - x_2} x_1 & \text{otherwise} \end{cases} \quad (6.5)$$

and the distance from the point $P = (x_P, y_P)$ to the line k equals to:

$$d(P, k) = \frac{|Ax_P + By_P + C|}{\sqrt{A^2 + B^2}}. \quad (6.6)$$

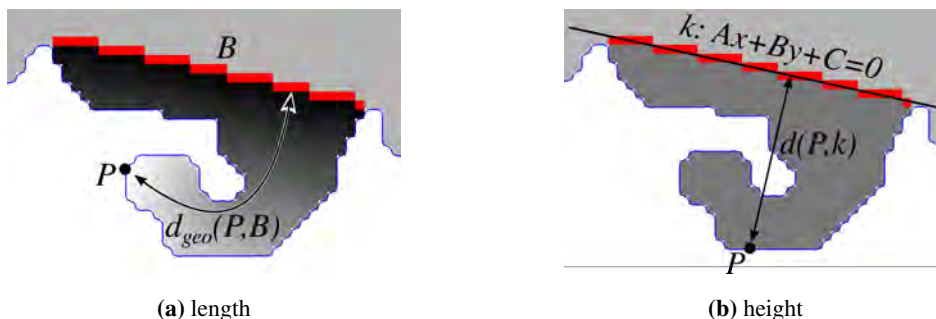


Figure 6.23. Measuring morphometry of a rete: (a) the rete length corresponds roughly to the maximum value of the geodesic distance transform of the rete region with seeds located at its base, (b) the rete height is the maximum distance of a rete boundary point P from the straight line $k : Ax + By + C = 0$ containing the rete base (the distance is computed for all points on the rete boundary).

6.2. Results

6.2.1. Evaluation metrics

We were interested mainly in answering the following three questions regarding performance of our algorithm: (1) how many retes are detected in the correct location (compared to the total number of retes), (2) how extreme is the over-detection problem, and (3) how closely the morphometry of automatically segmented retes matches the morphometry of manually segmented ones. The nature of our research problem required some fine-tuning of the popular evaluation metrics.

The fundamental issue is that retes bases, both manually and automatically segmented, are actually segments in the same 2D (high resolution) image space and thus it is virtually impossible that any pair of a manually and an automatically segmented rete bases will perfectly overlap – in most cases the matching endpoints of these bases will be located at least a few pixels apart. Therefore, we formulated the relaxed rete matching criterion as follows: in order to treat a rete R_1 as matching a rete R_2 the distance between the centers of their bases must be at most $l_2/2$, where l_2 is the length of R_2 's base.

Let us define the set of automatically segmented retes as \mathbf{R}_A and the set of manually segmented retes as \mathbf{R}_M . If either: (1) $R_A^i \in \mathbf{R}_A$ is matched by exactly one $R_M^j \in \mathbf{R}_M$ and R_M^j is matched only by R_A^i , or (2) $R_A^i \in \mathbf{R}_A$ is matched by exactly one $R_M^j \in \mathbf{R}_M$ and R_M^j is not matched by any rete from \mathbf{R}_A , or (3) $R_M^j \in \mathbf{R}_M$ is matched by exactly one $R_A^i \in \mathbf{R}_A$ and R_A^i is not matched by any rete from \mathbf{R}_M , then we will describe such two retes as “matched exclusively”. If a given manually segmented rete is matched by multiple automatically segmented retes, it implies that those automatically segmented retes should have been merged. Analogously, if a given automatically segmented rete is matched by multiple manually segmented retes, it implies that the automatically segmented rete should have be split.

The performance of our method was measured using two variations on the Jaccard index, a statistic used for comparing the similarity and diversity of finite sample sets. The Jaccard index is defined as the size of the intersection divided by the size of the union of the sample sets A and B :

$$J(A, B) = \begin{cases} \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} & \text{if } A \cup B \neq \emptyset \\ 1 & \text{otherwise} \end{cases} \quad (6.7)$$

and in our case the two sets are: (1) a set of manually segmented retes – the ground truth (denoted as MAN), and (2) a set of automatically segmented retes (AUT). In a perfect case, when all ground truth retes are matched and there are no automatically segmented retes without a match, the Jaccard index equals 1; otherwise its value is in the range $[0, 1)$. In the first variation of the Jaccard index, the restrictive one, the intersection set consists only of retes matching exclusively:

$$J_{\text{strict}} = \frac{\# \text{ true positive MANs}}{\# \text{ MANs} + \# \text{ false positive AUTs}} \quad (6.8)$$

In the relaxed one, automatically segmented retes matching non-exclusively (i.e., both groups of retes to be merged and retes to be split) also belong to the intersection set:

$$J_{\text{relaxed}} = \frac{\# \text{ MANs} - \# \text{ false negative MANs}}{\# \text{ MANs} + \# \text{ false positive AUTs}} \quad (6.9)$$

The relaxed variant is intended to reflect the fact that in many cases a pathologist arbitrary decides whether to treat a given epidermal projection as two individual retes (although partially joint) or as one rete with two tips – whereas both options are admissible (Fig. 6.24).

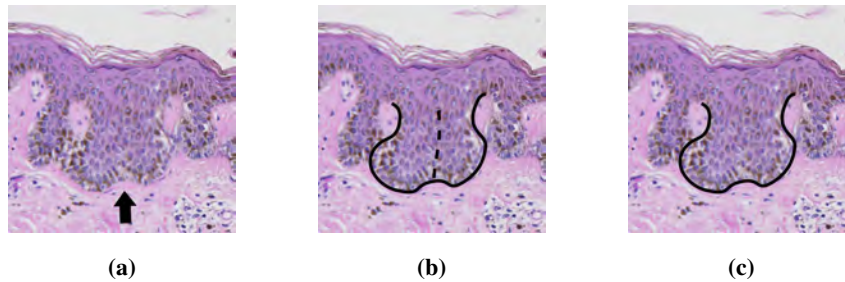


Figure 6.24. The rete splitting dilemma: the central region in (a) can be considered as either two joint retes as in (b) or as a single rete with two tips as in (c).

In order to increase the robustness of the above-mentioned metrics, prior to their computation we fine-tune the MAN and AUT sets in the following ways:

- Since edges of a skin specimen are often deformed due to mechanical slicing (as described in Section 2.2.4), all retes having roots located at most $130\ \mu\text{m}$ from main axis endpoints are excluded from both sets.
- All retes for which either their width w is too small ($w < w_{\min}$) or their length l is too small ($l < l_{\min}$) are excluded from both sets (we set $w_{\min} = 12\ \mu\text{m}$ and $l_{\min} = 25\ \mu\text{m}$, as for smaller projections even experienced pathologists decide whether to treat it as a rete or just as a slicing artifact in an arbitrary way).
- In case of the AUT set, all retes for which their length-to-width ratio is less than 0.5 are excluded.

However, that automatically segmented retes adjacent to holes in epidermis base are exempt from all those three rules (as each such a hole is always a sign of bridging between retes).

For retes matching exclusively we computed standard statistical measures describing the distribution of relative errors for both lengths of rete bases and lengths of retes themselves. For the true value of a quantity, x , and the inferred value, \hat{x} , the relative error is defined by

$$\delta x = \frac{\hat{x} - x}{x}. \quad (6.10)$$

6.2.2. Performance

The presented method has been evaluated on 25 skin whole slide images (WSIs) from UJCM-SS dataset, which included cases of lentigo (9), dysplastic nevus (11), and melanoma (5). The input to the algorithm consisted of manually segmented epidermis masks. There were 992 manually segmented retes in the whole dataset, whereas our method detected 894 retes: 825 were detected correctly (83%), 147 manually segmented retes were not detected (false negatives), in 20 cases either one manually segmented rete was assigned multiple automatically segmented retes or an automatically segmented retes should be split/merged, and 42 detected retes were actually false positives.

The distribution of retes as well as individual error types with respect to rete’s length and width is shown in Figure 6.25. The bulk of “should split/merge” errors are related to micro-retes: 62% of all such cases were related to retes both narrower than 60 μm and shorter than 40 μm whereas such tiny retes constitutes only 15% of all retes.

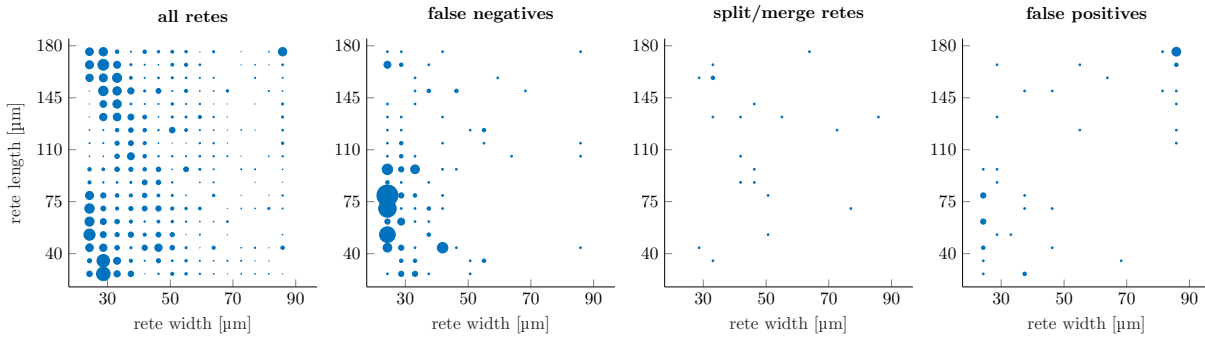


Figure 6.25. The distribution of retes with respect to their length and width (the marker scale for all error plots is the same, yet for the sake of clarity it is larger than the scale for the “all retes” plot)

The histogram of the proportion of incorrectly detected retes with respect to the rete length is shown in Figure 6.26. This proportion decreases rapidly as the length of a rete increases and the trend stabilize at 10–17% for retes longer than 40 μm (which constituted roughly 80% of all retes). The highest proportion of misdetections was observed among retes shorter than 35 μm, where nearly 65% of retes were missed (false negatives). The high proportion of false positives among retes longer than 90 μm (15 cases) may be explained by the misclassification of skin appendages (particularly: hair follicles) as retes. Even though the tissue structure of hairs and retes differ, from the morphometric point of view they are similar (Fig. 6.27). To the best of our knowledge no automatic method for the segmentation of skin appendages in histopathological images has been proposed so far.

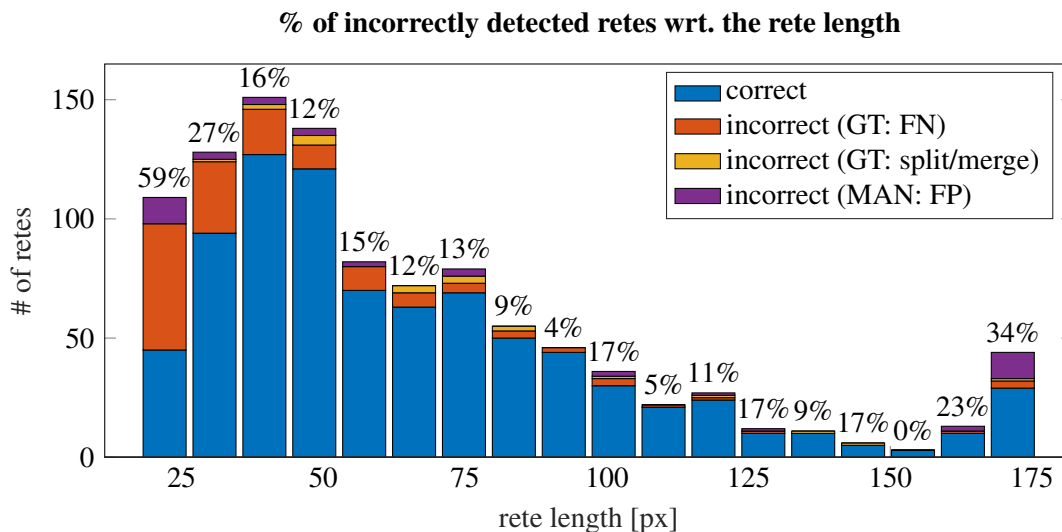


Figure 6.26. A histogram of the proportion of misdetections with respect to the rete length (the first bin also includes retes shorter than 25 μm and the last bin also includes all retes longer than 175 μm)

Figure 6.28 presents boxplots of J_{strict} and $J_{relaxed}$ indexes computed on the set of 25 images. The median and interquartile range (IQR) for J_{strict} were 0.80 and 0.10, whereas for $J_{relaxed}$ they equaled

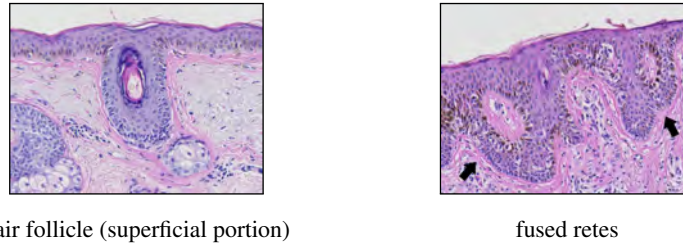


Figure 6.27. Morphometric similarity between skin appendages (e.g., hair follicles) and fused rete ridges

0.81 and 0.11, respectively. Both the strict and the relaxed version had the lower adjacent value (LAV) of nearly 0.70. The indexes computed for the aggregated data (i.e., from all slides altogether) were $J_{\text{strict}} = 0.798$ and $J_{\text{relaxed}} = 0.817$, respectively.

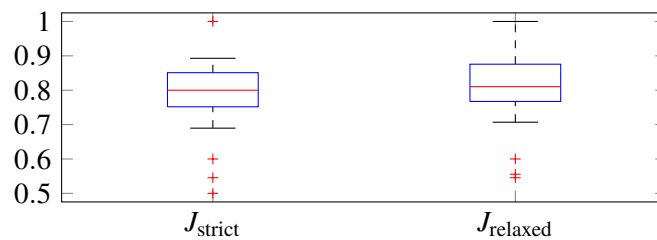


Figure 6.28. Boxplots of J_{strict} and J_{relaxed} indexes computed on the set of 25 images

The histograms of relative errors for estimating rete width and length are shown in Figure 6.29. The mean absolute value of relative errors ($|\delta x|$) was 20% for the width and 16% for the length. Both distributions of relative errors were right-skewed – for width errors skewness was moderate (0.79), whereas for length errors it was high (3.42) – suggesting that our method rather overestimates than underestimates both width and length of retes. Making a similar analysis for rete height is pointless, since even a tiny change in the orientation of a rete base has a considerable impact on the height estimate.

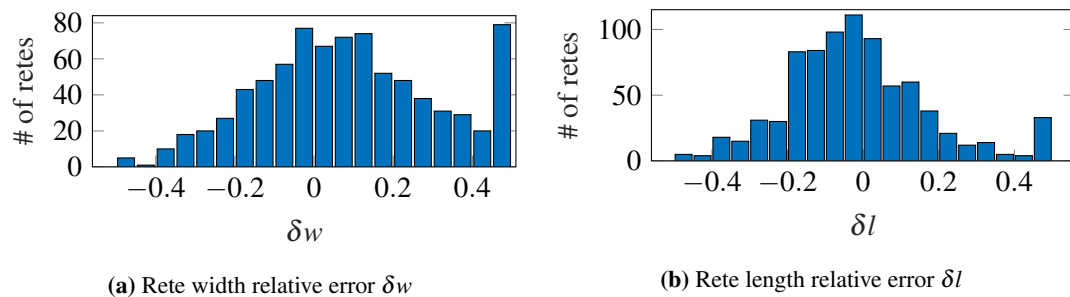


Figure 6.29. Histograms of relative errors for: (a) the rete width, and (b) the rete length. In each histogram the first bin also includes errors smaller than -0.50 and the last bin also includes errors larger than 0.50 .

Additionally, we evaluated the performance of the rete segmentation method on a set of 75 images of automatically segmented epidermis regions (obtained using the method described in Chapter 5).

There were 3709 manually segmented retes in the whole image set, whereas our method detected 2613 retes: 1571 were detected correctly (42%), 1932 manually segmented retes were not detected (false negatives), in 206 cases either one manually segmented rete was assigned multiple automatically segmented

retes or an automatically segmented retes should be split/merged, and 874 detected retes were actually false positives.

Figure 6.30 presents boxplots of J_{strict} and J_{relaxed} indexes computed on this image set. The median and interquartile range (IQR) for J_{strict} were 0.38 and 0.18, whereas for J_{relaxed} they equaled 0.40 and 0.19, respectively. Both the strict and the relaxed version had the lower adjacent value (LAV) of nearly 0.03. The indexes computed for the aggregated data (i.e., from all slides altogether) were $J_{\text{strict}} = 0.354$ and $J_{\text{relaxed}} = 0.368$, respectively.

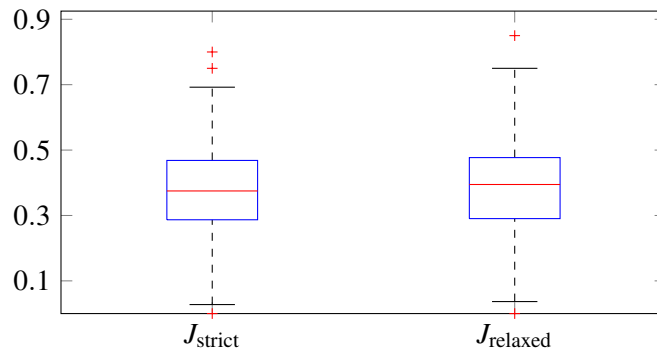


Figure 6.30. Boxplots of J_{strict} and J_{relaxed} indexes computed on the set of 75 images of automatically segmented epidermis

Even though this time the results might appear unsatisfactory, in Chapter 7 we will prove that they still constitute meaningful input to train a well-performing lesion classifier model. Such low values of the true positive rate, and Jaccard indexes are a consequence of the nature of our epidermis segmentation method: it produces a more jagged boundary of the epidermal region (i.e., with more micro-protrusions), which causes the boundary nodes to be located in slightly different places. In case of narrow retes those tiny discrepancies often have serious consequences: even though a given automatically segmented rete (base) may be located “more or less” close to a manually segmented rete (base), so that a human observer would definitely treat them as matching retes, in fact the distance between their centers would be too large for the automatic rete matching criterion. However, most of the features described in Section 7.1 utilize either only the information about the sole fact of rete’s existence (e.g., the density of all retes) or are based on the analysis of a long section of epidermis (e.g., the median thickness of epidermal base) and thus are robust to the aforementioned slight discrepancies in the location and/or morphometry of some retes.

6.3. Conclusions and discussion

We proposed a method which automatically measures the morphometry of a segmented epidermis – it determines the location, width, length, and height of individual rete ridges – and tested it on two datasets consisting of segmented epidermis masks prepared from skin WSIs of common melanocytic lesions: the first included 25 images of manually segmented epidermal regions, whereas the second included 75 images of automatically segmented epidermal regions (using the algorithm described in Chapter 5).

Experimental results are promising: for the first dataset, in total 825 out of 992 (83%) manually segmented retes were detected correctly and the “strict” Jaccard similarity coefficient for the task of detecting rete ridges was 0.798. The main sources of errors were misdetection of micro-retes and treating skin appendages as retes. The relative error of width and height estimates is on average $\pm 20\%$ and $\pm 16\%$, respectively. Our method tends to overestimate rather than underestimate values of these morphometric features. Even though statistical results obtained for the second dataset (including automatically segmented epidermal regions) were much worse, the actual measurements of rete morphometry were good enough to obtain high classification performance, as described in Chapter 7. There are three main known limitations of our method (shown in Figure 6.31):

1. If a group of retes fused near the epidermal plate is too wide, the method fails to determine the rete base of such a group.
2. Long retes at the end of the lesion are erroneously treated as part of the main axis (actually, it is problematic to distinguish a short “tip” of the epidermal plate from a rete, as typically such a rete is close to the lesion boundary).
3. Long retes not growing nearly perpendicularly into dermis (but rather joined with the epidermal plate at multiple places along their boundary) will often be split by the algorithm as if they were series of short retes fused together.

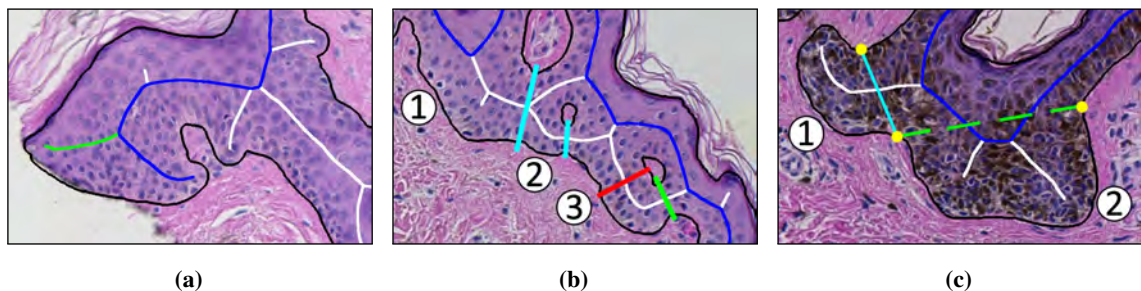


Figure 6.31. The three main shortcomings of our method: (a) long retes at the end of the lesion are erroneously treated as part of the main axis (the actual main axis and its proper ending is marked blue and green, respectively); (b) long retes not growing nearly perpendicularly into dermis, but rather joined with the epidermal plate at multiple places along their boundary (such as the rete #3), will often be split by the algorithm as if they were series of short retes fused together (the actual and the proper split is marked red and green, respectively); (c) if a group of retes fused near the epidermal plate is too wide, the method fails to determine the rete base of such a group (as it is the case for the group #2).

To the best of our knowledge it is the first method to measure morphometric features of individual boundary projections as well as the first one to show its application to analyze the epidermal morphometry in skin histopathological images. However, the application of our method is not limited to the field of dermatopathology – after minor adjustments it can be used for pathological diagnostics of organs covered with the mucous membrane or in dermatoscopy (to assess border irregularity of melanocytic lesions [132], as a jagged lesion border is the hallmark of melanoma).

In our study, the results obtained with this algorithm are later used to compute values of more complex indexes describing the epidermal morphometry (such as variations in epidermal hyperplasia and changes

in elongation of rete ridges along the lesion) which constitute a feature vector for skin lesion classification models, as described in Chapter 7.

6.4. Summary

In this chapter we presented an algorithm for automatically measuring the basic morphometry of individual retes in a segmented epidermis. It firstly determines the route of the epidermis main axis by searching for the best paths (in terms of criteria based on the path length and distance of its endpoints from the epidermis “outer” edge) in graphs spanned on the skeleton of the epidermal region. Then, it approximates the location of retes (i.e., their roots) along that axis by finding those skeleton nodes in which edges pointing towards the epidermis “outer” edge branch off. The candidates for endpoints of rete bases are identified by analyzing the curvature of epidermis boundary; the actual endpoints are then matched with rete roots according to geometric criteria, such as: the relative orientation of a “projection root–candidate node” vector with respect to the orientation of a given projection, and the distance between a node and the root. As retes often either contain multiple tips or are partially joined with their neighboring retes, a post-processing step (merging and/or spiting) is required. Finally, the morphometry (width, length and height) of individual retes is measured. We proposed metrics useful for the evaluation of the proposed algorithm, provided the segmentation results, and verified it against the ground truth prepared by an experienced dermatopathologist.

7. Feature extraction and classification

To the best of our knowledge, in the literature there are only a few works about automatic diagnosis of melanocytic lesions seen in H&E-stained skin whole slide biopsy images. Xu et al. [133] proposed a system for automated classification of skin melanocytic tumors based on the method for the melanocytes detection in [134] and the method for the epidermis segmentation in [113]. It consists of the following steps: segmentation of both epidermis and dermis, computation of a set of features related to nuclear morphology and spatial distribution of melanocytes, lesion classification based on the feature vector using a multi-class support vector machine. Noroozi and Zakerolhosseini [135] proposed an automated algorithm capable of differentiating squamous cell carcinoma in situ from actinic keratosis. Its initial step is epidermis segmentation, followed by the removal of stratum corneum; then, epidermis main axis is specified using paths in the skeleton of the epidermal region, and the granular layer is removed via connected components analysis; finally, lesion classification is performed based on intensity profiles along lines perpendicular to the epidermis main axis. Olsen et al. [136] used deep neural networks to classify lesion images of basal cell carcinomas, dermal nevi, and seborrheic keratoses.

None of the aforementioned methods actually considers the morphometry of rete ridges and epidermal component as a whole, although uniformity and symmetry of both elongation and thickening of rete ridges as well as uniformity and symmetry of epidermal hyperplasia are among most substantial diagnostic criteria for skin melanocytic lesions [16, 17]. Therefore, we consider the issue of automatic diagnosis using histopathological images as not yet sufficiently addressed. Since histopathological image analysis is the gold standard for diagnosing and grading skin tissue malignancies, our method will provide a valuable contribution towards automating this procedure.

7.1. Characteristic features for epidermal morphometry

Based on the clinical experience of dermatopathologists [16, 17] we proposed the following indexes describing the morphometry of the epidermal component, which can be used to classify skin melanocytic lesions:

- LESION_LEN** the lesion length L_π (the length of the epidermis main axis)
- RHO_RETE_ALL** the density of all retes: n_{retes}/L_π

- RHO_RETE_LONG** the density of “long” retes: the number of those retes for which the ratio of their height to the thickness of the epidermal base at their roots is greater than 0.75, divided by L_π
- THICKNESS_MED** the median thickness of the epidermal base
- N_HYPERPLASTIC** the number of hyperplastic sections of the epidermis, i.e., sections which are 2 times thicker than the median thickness of epidermis base and they are at least 90 μm wide (such a threshold was adopted since in some cases the rete detection algorithm erroneously places the rete base too far from the actual boundary of the epidermal base, whereas nearly all retes are at most 90 μm wide, as seen in Figure 6.25)
- RHO_HYPERPLASTIC** the density of hyperplastic sections: the ratio of the length of all hyperplastic sections to the lesion length L_π
- RHO_RETE_LOSS** the proportion of the length of sections with loss of rete ridges (i.e., sections where for at least 200 μm there is not a single long rete) to the total lesion length L_π (such a threshold was adopted since nearly all retes are at most 90 μm wide, as seen in Figure 6.25, but individual retes are typically separated one from another by dermal papillae which are often up to 50 μm wide)

The lesion length L_π is measured as follows:

$$L_\pi = R_l \sum_{i=1}^{|\pi_m|-1} \sqrt{(P_i^x - P_{i+1}^x)^2 + (P_i^y - P_{i+1}^y)^2}, \quad (7.1)$$

i.e., each pair of either horizontal or vertical neighbors is counted as 1 and each pair of diagonal neighbors is counted as $\sqrt{2}$.

The procedure used to measure the thickness of epidermis as a function of the lesion run length consists of the following steps. Firstly, at each point of the epidermis main axis two rays are cast perpendicularly to that axis – but in opposite directions – and for each of these rays the distance to the intersection with the boundary of the epidermis base region is determined; those distances are then added up. Then, for each rete the thickness along the section of the main axis within the neighborhood of that rete is linearly interpolated between its endpoints (i.e., for each rete we consider the section consisting of pixels lying at most $\pm l/2$ from the center of the rete root, where l is the length of the rete base). Finally, in order to reduce artifacts, the signal is smoothed using a moving average filtering with the window of size $w = 5 \text{ px}$.

Figure 7.1 shows the process of feature extraction for computing the above-mentioned morphometric indexes.

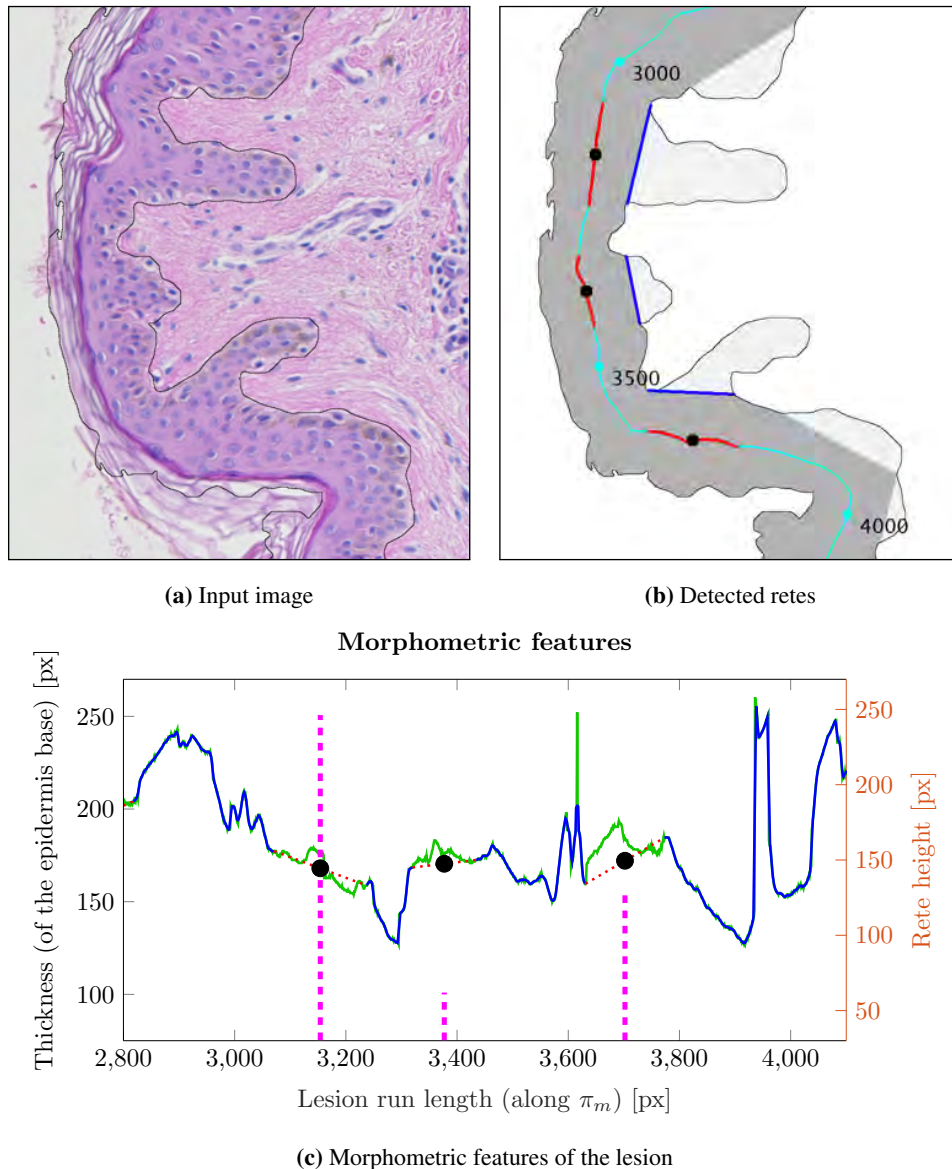


Figure 7.1. The process of feature extraction for computing morphometric indexes: (a) an extract from the input WSI image (the segmented epidermis is marked with a black line); (b) epidermal morphometry as detected by our algorithm: epidermal base (dark gray), rete roots (light gray), epidermis main axis (the cyan line, numbers denote the main axis run length), rete bases (blue lines), rete roots (black dots), and “ranges” of retes along the epidermis main axis (red sections of the main axis); (c) a plot depicting basic morphometric features from (b): the raw thickness function (the solid green line), the smoothed and interpolated thickness function (the solid blue line and dotted red lines), and the location and height of retes (dashed purple lines).

7.2. Classification

In our study we compared the performance of the following classifiers: multinomial logistic regression, naïve Bayes classifier, classification tree, and support vector machine (SVM). In particular, the classification tree and multinomial logistic regression models were included as they yield interpretable

results, which can then help propose specific diagnostic guidelines for clinicians (e.g., which features were most important). Neural networks, although being often used for complex tasks in medical image processing [137, 138], require a large training dataset in order to make accurate predictions and thus were not considered in this study.

7.2.1. Multinomial logistic regression

The multinomial logistic regression is a classification model used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables, i.e., we have a training set $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ of m labeled examples, where the input features are $x^{(i)} \in \mathbb{R}^n$, and observation labels $y^{(i)} \in \{1, \dots, K\}$, where K is the number of classes. Given a test input x , the hypothesis $h_\theta(x)$ estimates the probability that $P(y = k | x)$ for each value of $k = 1, \dots, K$ (it uses a standard softmax function to turn the scores $\theta^\top x$ outputted by the model into probabilities) [139]:

$$h_\theta(x) = \begin{bmatrix} P(y = 1 | x; \theta) \\ P(y = 2 | x; \theta) \\ \vdots \\ P(y = K | x; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^K \exp(\theta^{(j)\top} x)} \begin{bmatrix} \exp(\theta^{(1)\top} x) \\ \exp(\theta^{(2)\top} x) \\ \vdots \\ \exp(\theta^{(K)\top} x) \end{bmatrix} \quad (7.2)$$

The accuracy of the model for a given observation $x^{(i)}$ from the training set is assessed by measuring the distance D between the two probability vectors: the one containing class probabilities outputted by the classifier $S = h_\theta(x^{(i)})$ and the one-hot encoded vector L that corresponds to the true label (in one-hot encoding, the target vector for an observation belonging to class C_k is a binary vector of size K with all elements zero except for element k , which equals one). In our case, the cross-entropy function is used [139]:

$$D(S, L) = - \sum_{j=1}^K L_j \log(S_j) \quad (7.3)$$

To transform the multinomial classification problem into a proper optimization problem the training loss is defined as the cross-entropy loss averaged over the entire training dataset.

The multinomial logistic regression model was fitted using the Stochastic Average Gradient (SAG) method [140] with L2 regularization (and regularization parameter $C = 1$) and the tolerance for stopping criteria $\varepsilon = 0.1$. The training algorithm used the cross-entropy loss. The usage of the SAG solver required that the training would be performed on the standardized data, i.e., prior to the training each column of the predictor data was centered and scaled by the weighted column mean and standard deviation, respectively. All observations were equally weighted.

7.2.2. Naïve Bayes classifier

Naïve Bayes classifiers are a set of supervised learning algorithms based on applying Bayes' theorem with the "naïve" assumption of conditional independence between every pair of features given the value

of the class variable [141]. Bayes' theorem states the following relationship, given class variable y and dependent feature vector $x \in \mathbb{R}^n$:

$$P(y | x) = \frac{P(y)P(x | y)}{P(x)} \quad (7.4)$$

Using the naïve conditional independence assumption that

$$P(x_i | y; x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y),$$

for all i , the relationship in (7.4) is simplified to

$$P(y | x) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x)}. \quad (7.5)$$

As $P(x)$ is constant given the input, the following classification rule can be applied:

$$\begin{aligned} P(y | x) &\propto P(y) \prod_{i=1}^n P(x_i | y) \\ &\Downarrow \\ \hat{y} &= \arg \max_y P(y) \prod_{i=1}^n P(x_i | y), \end{aligned}$$

and the Maximum A Posteriori (MAP) estimation can be used to estimate $P(y)$ and $P(x_i | y)$; the former is then the relative frequency of class y in the training set. The different naïve Bayes classifiers differ mainly by the assumptions they make regarding the distribution of $P(x_i | y)$.

In our experiment, the naïve Bayes classifier was fitted assuming predictors have Gaussian distribution and class prior probabilities equal the frequency of occurrence of each class in the dataset. All observations were equally weighted.

7.2.3. Classification tree

A classification tree uses a decision tree (as a predictive model) to go from remarks about an observation (represented in the branches) to conclusions about the observation's class label (represented in the leaves) – branches represent conjunctions of features that lead to those class labels.

The basic idea of decision tree growing is to choose one among all the possible splits at each node so that the resulting child nodes are the “purest”. One of the most popular tree-growing algorithm is the CART algorithm based on “Classification and Regression Trees” by Breiman et al. [142], which considers only univariate splits (i.e., each split depends on the value of only one predictor variable).

To grow a CART tree for continuous features, the following steps are repeatedly used on each node starting from the root [142]:

1. Find each feature's best split.

For each feature, sort its values in ascending order and go through them (starting from top) to examine each candidate split point v : if the observed feature value is less than or equal v , the case goes to the left child node, otherwise, it goes to the right. The best split point is the one that maximize

the splitting criterion the most when the node is split according to it (an example of a splitting criterion is provided later).

2. Find the node's best split.

Among the best splits found in step 1, choose the one that maximizes the splitting criterion.

3. If the stopping rules are not satisfied, split the node using its best split found in step 2.

A commonly used splitting criterion is the *Gini index*, which measures the total variance across the K classes and is defined as:

$$\text{Gini}(p) = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (7.6)$$

where \hat{p}_{mk} represents the proportion of training observations in the m th region (leaf node) that are from the k th class [143].

In the process of growing a CART tree the following stopping rules are used:

- If a node becomes pure (i.e., it consists only of cases belonging to the same class), the node will not be split.
- If all cases in a node have identical feature vectors, the node will not be split.
- If either the size of a node is less than the user-specified minimum node size value or the split of that node results in a child node whose node size is less than that value, the node will not be split.
- If for the best split of a node the improvement is smaller than the user-specified minimum improvement, the node will not be split.
- The tree depth cannot exceed the user-specified maximum tree depth limit value.

To reduce overfitting (and thus improves predictive accuracy), decision trees are often pruned, e.g., by discarding the n deepest levels.

The process of predicting a class label of an observation is sequential and hierarchical, and goes as follows:

1. Start from the root node of the tree.
2. Compare the feature split value in the node with the observed value of that feature: if the observed value is less than or equal the split value, follow the left branch, otherwise, follow the right branch.
3. Repeat the step 2 until reaching a leaf node.

The decision tree classifier was trained using the standard Classification and Regression Trees (CART) algorithm (i.e., it selects the split predictor that maximizes the split-criterion gain over all possible splits of all predictors – in this case, the Gini's Diversity Index) [142]. To prevent overfitting, model hyperparameters were obtained via MATLAB's hyperparameters optimization, which attempts to minimize the cross-validation loss (error) by varying the parameters: the minimum number of leaf node observations was set to 3, the minimum number of branch node observations was set to 10, for each split all features were selected, and the resulting tree was pruned to the 1st level (the value was chosen based on the stratified cross-validated classification error criterion). Additionally, leaves that originated from the same parent node, and that yielded a sum of risk values greater or equal to the risk associated with the parent node, were merged. All observations were equally weighted.

7.2.4. Support vector machine

The (binary) support vector machine (SVM) model represents the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they lie [141].

Linearly separable data admits infinitely many decision boundaries that separate the classes, but intuitively some of these are better than others. If we take the score of example \mathbf{x} as $\hat{s}(x) = \mathbf{w}\mathbf{x} - b$, where \mathbf{w} is a normal vector of the hyperplane and b is the distance between the decision boundary and the nearest training instances (at least one of each class) as measured along \mathbf{w} , then a true positive \mathbf{x}_i satisfies $\mathbf{w}\mathbf{x}_i - b > 0$ and a true negative satisfies $\mathbf{x}_j - (\mathbf{w}\mathbf{x}_j - b) > 0$. The $\frac{|b|}{\|\mathbf{w}\|}$ proportion is called the *margin* and the goal of the SVM algorithm is to find a decision boundary maximizing that margin.

Maximizing the margin corresponds to minimizing $\|\mathbf{w}\|$ (provided that none of the training points fall inside the margin), leading to a quadratic, constrained optimization problem:

$$\mathbf{w}^*, b^* = \underset{\mathbf{w}, b}{\operatorname{arg\,min}} \|\mathbf{w}\|^2 \quad \text{subject to} \quad \forall_i : y_i(\mathbf{x}_i\mathbf{w} + b) \geq 1 \quad (7.7)$$

For the classification problem, only training examples satisfying the relationship (7.7) (i.e., only training examples nearest to the decision boundary) are relevant – they are called *support vectors* (Fig. 7.2).

The equation (7.7) can be used only for linearly separable datasets. In order to be able to deal with the data that are not linearly separable, a series of slack variables ξ_i is introduced (one for each example $i = 1, 2, \dots, n$). The constraint then takes the following form:

$$\forall_i : y_i(\mathbf{x}_i\mathbf{w} + b) \geq 1 - \xi_i \quad \text{subject to} \quad \forall_i : \xi_i \geq 0. \quad (7.8)$$

In order to minimize $\|\mathbf{w}\|^2$ the method of Lagrange multipliers is used. Adding the constraints with multipliers α_i for each training example gives the Lagrange function:

$$\Lambda(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j \quad \text{subject to} \quad \forall_i : \alpha_i \geq 0 \quad \text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (7.9)$$

To allow the user to define the trade-off between margin maximization and slack variable minimization, a parameter C is introduced to the equation (7.9):

$$\Lambda_{\text{nonlinear}}(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j \quad (7.10)$$

subject to

$$\forall_i : 0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

The generalization of the above equations on nonlinear decision function is done using *kernels*. It turns out that the solution to the linear problem involves only the inner products of the observations: $\mathbf{x}_i \mathbf{x}_j$. If the data were mapped into another Euclidean space H (of any number of dimensions) using some transformation Φ , then the learning algorithm would depend on the original (untransformed) data

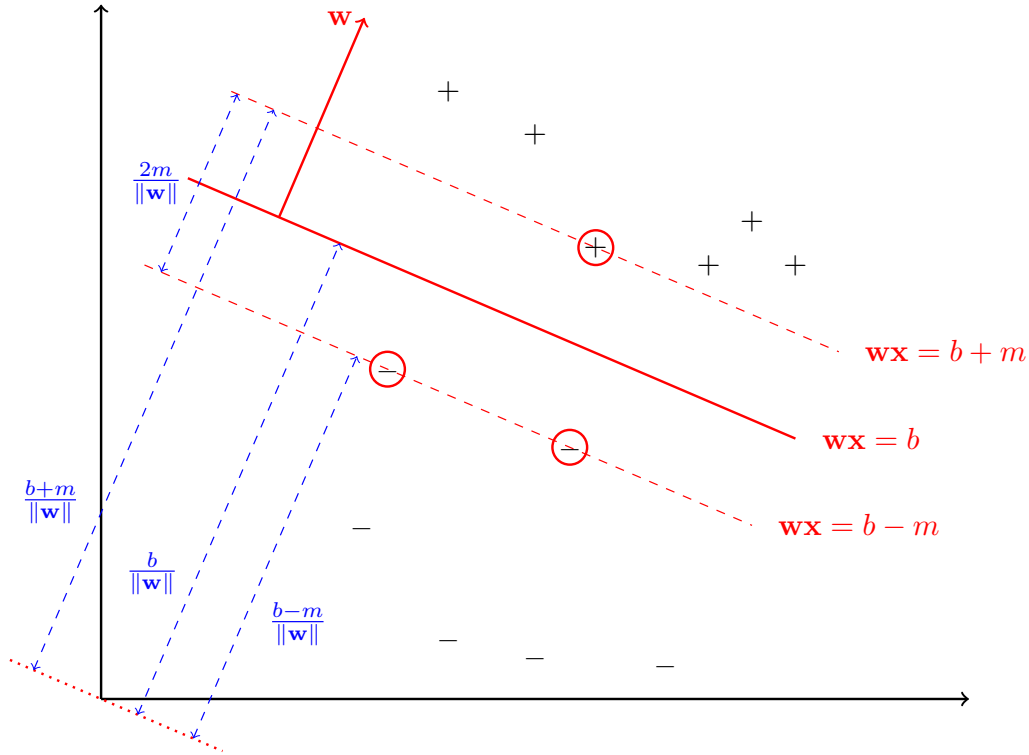


Figure 7.2. The idea of a (linear) SVM. The circled data points are the support vectors, which are the training examples nearest to the decision boundary. The SVM finds the decision boundary that maximizes the margin $\frac{b}{\|w\|}$.

via the dot product included in H , i.e., via $\Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$. By defining a kernel function K as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)$ it is enough to use that function in (7.10) and the knowledge of the exact form of Φ is inessential:

$$\Lambda_k(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (7.11)$$

subject to

$$\forall_i : 0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

Such a reformulation allows to map the data from lower to higher dimensional space without the need of performing additional computations. Some of the most common kernel functions used in SVMs are summarized in Table 7.1.

Due to various complexities, a direct solution of multiclass problems using a single SVM formulation is usually avoided; instead, a combination of several binary SVM classifiers is used to solve a given multiclass problem [144]. A common strategy is the “winner-takes-all” (WTA) strategy, in which K binary classifiers are constructed: the i th classifier output function ρ_i is trained taking the examples from the \mathcal{C}_i class as positive and the examples from all other classes as negative – a new example x is assigned to the class with the largest value of ρ_i . Other strategies include among others max-wins voting strategy [145], pairwise coupling strategy [146], and directed acyclic graph method [147].

Table 7.1. The selected types of kernel functions

Kernel type	Kernel formula	Limitations
linear	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j$	
polynomial	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \mathbf{x}_j + r)^d$	$\gamma > 0$
radial basis function (RBF)	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right)$	
sigmoidal	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i \mathbf{x}_j + r)$	$\gamma > 0$

To classify using the SVM model, a multiclass error-correcting output codes (ECOC) model was trained utilizing the one-vs-all coding for SVMs with the RBF kernel (of scale $\gamma = 1$) and the box constraint $C = 1$. The RBF kernel has been used since the classification problem stated in this study consists of sets of observations which cannot be separated linearly (other benefits of using the RBF kernel are described in [148]). The the WTA strategy was used as for small datasets it yields results comparable to other strategies which in turn (for larger datasets) typically outperform the WTA strategy [144]. The training was performed on the standardized data. All observations were equally weighted. The class prior probabilities were set to the class relative frequencies in the labeling. The training was performed using the sequential minimal optimization (SMO) algorithm [149].

7.3. Results and discussion

To compare the performance of the selected classifiers we performed a 5-fold stratified cross-validation (CV) and repeated the procedure 100 times ensuring a unique partitioning in each repetition (due to a relatively small size of the dataset performing a cross-validation procedure only once would not yield representative results).

The k -fold CV procedure consists of the follows steps. Firstly, the dataset is randomly split into k groups. Then, for each unique group that group is taken as a test dataset, whereas observations from the remaining $k - 1$ groups are used to fit a model (the score obtained by evaluating the model on the test set is retained). Finally, the performance of the general model is summarized by averaging evaluation scores of k fitted models. By averaging over training sets one gets a sense of the variance of the learning algorithm (however, the training sets in cross-validation have considerable overlap and thus are not independent) [141].

We trained and compared classifier models in two groups. In the first one the cost of classifying a point into class j if its true class is i was defined as:

$$\text{Cost}_{\text{eq}}(i, j) = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{otherwise} \end{cases}. \quad (7.12)$$

However, since in dermatopathology the cost of misclassifying a malignant lesion as a benign one is more severe than the cost of misclassifying a benign lesion as a malignant one, in the second group we trained models using the following (mis)classification cost:

$$\text{Cost}_{\text{diff}}(i, j) = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \wedge i \neq \text{melanoma} \\ 2 & \text{if } i \neq j \wedge i = \text{melanoma} \end{cases} \quad (7.13)$$

Boxplots of mean classification errors for each classifier for Cost_{eq} are shown in Figure 7.3, whereas descriptive statistics of each error distribution are summarized in Table 7.2. Same plots and statistics for $\text{Cost}_{\text{diff}}$ are presented in Figure 7.4 and Table 7.3. In case of the logistic regression model it is not possible to directly include the non-uniform misclassification cost. For both cost matrices the SVM classifier scored lowest median of mean CV error rates. Increasing the “malignant as benign” misclassification cost resulted in higher median of mean CV error rates for the decision tree and naive Bayes classifiers, whereas for the SVM classifier the median decreased.

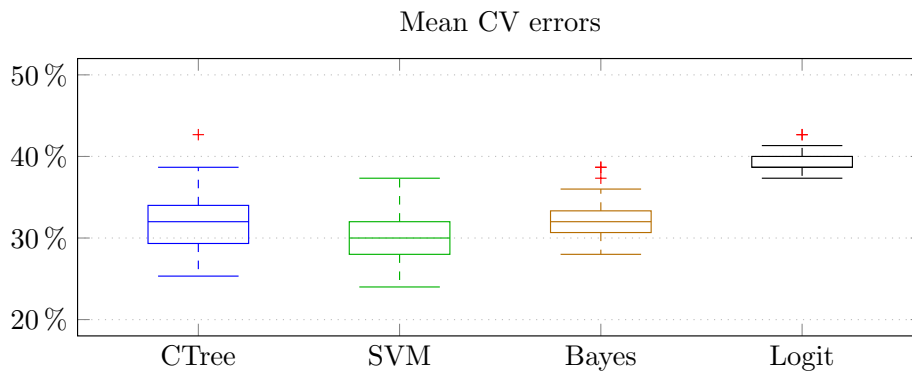


Figure 7.3. Boxplots of mean CV classification error rates for Cost_{eq} (outliers are marked with a “+” mark)

Table 7.2. Summary statistics of mean CV classification error rates for Cost_{eq}

Classifier	μ	σ	Q1	median	Q3
CTree	0.321	0.033	0.307	0.320	0.347
SVM	0.305	0.027	0.280	0.307	0.320
Bayes	0.323	0.023	0.307	0.320	0.333
Logit	0.391	0.013	0.387	0.387	0.400

In order to compare models fitted using different cost matrices we performed leave-out-one cross validation (LOOCV). The LOOCV procedure is simply a special case of k -fold CV for $k = n$, where n is the number of observations in the dataset. Typically, the LOOCV yields better approximation of the classification error than the k -fold cross validation (for $k < n$) [150, 151], especially for small datasets. Confusion matrices and LOOCV errors for a SVM and a decision tree are presented in Figures 7.5 and 7.6. In case of the SVM classifier changing the cost of a “malignant as benign” misclassification

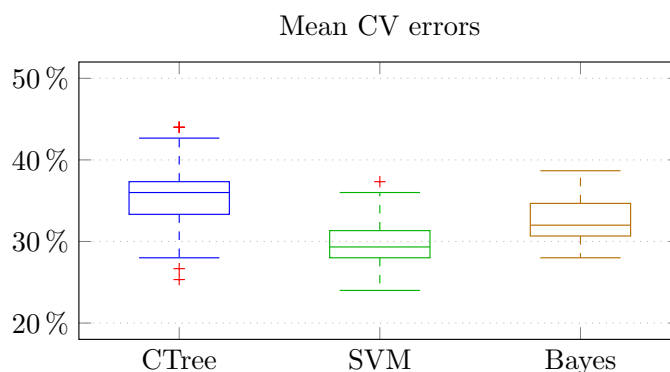


Figure 7.4. Boxplots of mean CV classification error rates for $\text{Cost}_{\text{diff}}$ (outliers are marked with a “+” mark)

Table 7.3. Summary statistics of mean CV classification error rates for $\text{Cost}_{\text{diff}}$

Classifier	μ	σ	Q1	median	Q3
CTree	0.354	0.038	0.333	0.360	0.373
SVM	0.300	0.029	0.280	0.293	0.313
Bayes	0.329	0.024	0.307	0.320	0.347

resulted in less errors of this type and in a lower LOOCV error rate, whereas in case of the decision tree classifier there was no change in the number of errors of this type and the LOOCV error rate actually increased.

		Predicted							
		JD	Lentigo	Melanoma					
Actual	JD	17	6	2	Actual	JD	19	3	3
	Lentigo	7	13	5		Lentigo	6	13	6
	Melanoma	1	2	22		Melanoma	0	1	24
LOOCV error: 30.7%					LOOCV error: 25.3%				
(a) Cost_{eq}					(b) $\text{Cost}_{\text{diff}}$				

Figure 7.5. Confusion matrices obtained for a SVM using leave-one-out cross-validation (LOOCV) for two different cost matrices: (a) Cost_{eq} , and (b) $\text{Cost}_{\text{diff}}$.

Since one of our research objectives was to formulate the basic histopathological criteria related to epidermal morphometry, we analyzed the results obtained using the decision tree classifier in details. Figure 7.7 shows a schema of the most accurate decision tree model fitted on the whole dataset using the Cost_{eq} misclassification cost and the same parameter values as described in 7.2, for which the LOOCV error equaled 29.3%. It can be seen clearly, that predictors and thresholds used for splits are concordant with the clinical practice: the lesion length depends mainly on how advanced that skin lesion is; since in

		Predicted					Predicted		
		JD	Lentigo	Melanoma			JD	Lentigo	Melanoma
Actual	JD	18	6	1	Actual	JD	20	5	0
	Lentigo	4	16	5		Lentigo	6	16	3
	Melanoma	3	3	19		Melanoma	3	3	19

LOOCV error: 29.3% LOOCV error: 26.7%

(a) $Cost_{eq}$ (b) $Cost_{diff}$

Figure 7.6. Confusion matrices obtained for a decision tree using leave-one-out cross-validation (LOOCV) for two different cost matrices: (a) $Cost_{eq}$, and (b) $Cost_{diff}$.

lentigo the epidermal component is effaced and there is a severe loss of rete ridges, the median thickness and the density of retes (regardless their length) is low; in junctional dysplastic nevi the epidermal architecture is generally preserved, i.e., there are numerous retes (both short and long ones) and no abnormal thinning of the epidermis. Unfortunately, it may be hard to fully apply the diagnostic schema proposed by the decision tree model as diagnostic guidelines for clinicians, as it is hard to measure the median thickness of the epidermal base with the naked eye.

Since the most important question when diagnosing a skin melanocytic lesion is whether it is benign or malignant, in Table 7.4 we summarized performance of each classifier for “one-vs-all” binary classification problems – the highest sensitivity for melanoma detection (96.0%) was scored by a SVM model trained using the $Cost_{diff}$ cost matrix.

Table 7.4. Accuracy, sensitivity and specificity of “one-vs-all” classifications

Index	SVM ($Cost_{eq}$)	SVM ($Cost_{diff}$)	CTree ($Cost_{eq}$)	CTree ($Cost_{diff}$)
Acc_{Ln}	72.0%	78.7%	76.0%	77.3%
Sen_{Ln}	52.0%	52.0%	64.0%	64.0%
Spe_{Ln}	82.0%	92.0%	82.0%	84.0%
Acc_{JD}	77.3%	84.0%	81.3%	81.3%
Sen_{JD}	64.0%	76.0%	72.0%	80.0%
Spe_{JD}	84.0%	88.0%	86.0%	82.0%
Acc_{Me}	86.7%	86.7%	84.0%	88.0%
Sen_{Me}	88.0%	96.0%	76.0%	76.0%
Spe_{Me}	86.0%	82.0%	88.0%	94.0%

Ln – lentigo, JD – junctional dysplastic nevus, Me – melanoma

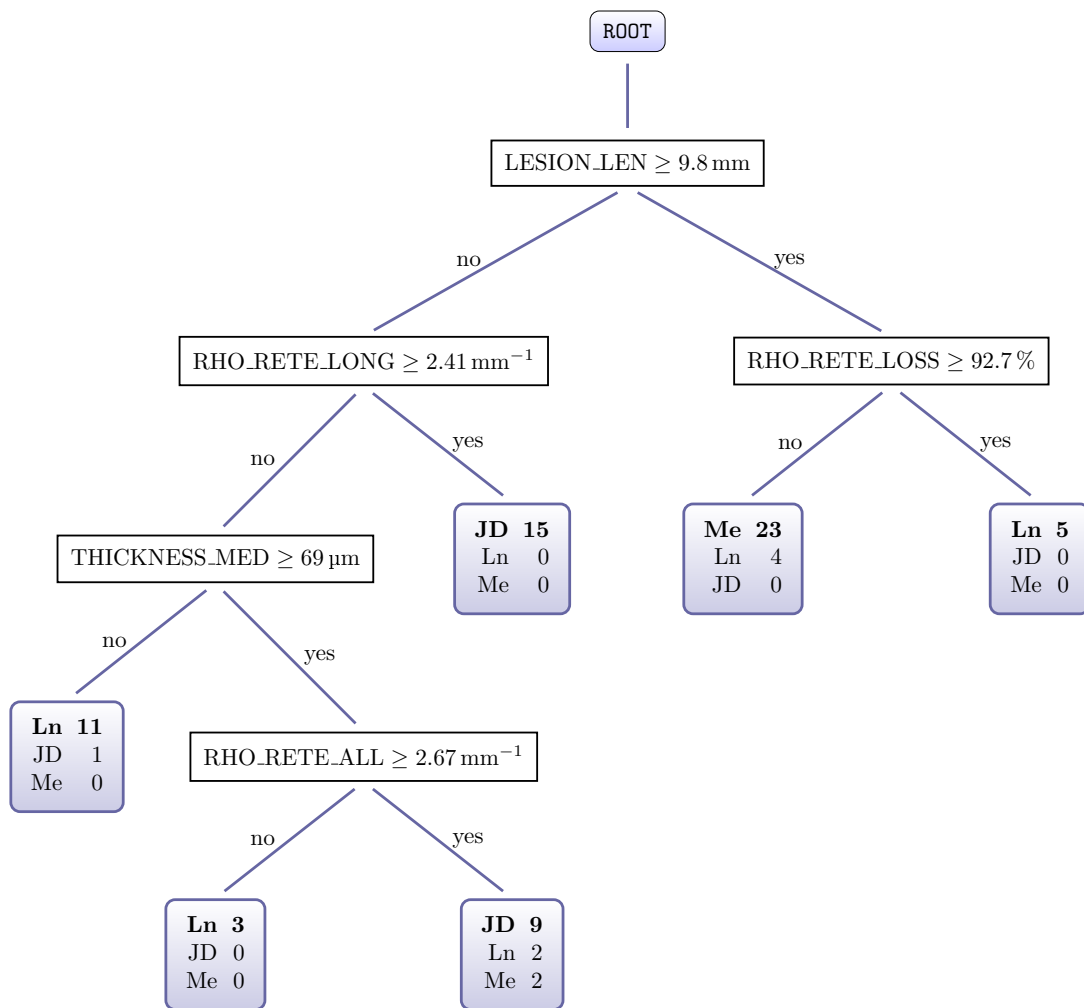


Figure 7.7. A schema of the decision tree model fitted on the whole dataset for $Cost_{eq}$. In each leaf node the majority class was marked bold, whereas cases belonging to other classes represent misclassified cases (Ln – lentigo, JD – junctional dysplastic nevus, Me – melanoma).

Most misclassification errors result from one of the following reasons:

- The epidermis segmentation method fails to remove thick stratum corneum from the resulting segmentation, especially when individual layers of corneocytes resemble puff pastry and are additionally blurred (due to the lack of focus when taking a photo of the lesion boundary region). Consequently, in some cases the median thickness of the epidermal base will be inflated.
- The rete segmentation method does not distinguish epidermal projections being part of skin appendages (mainly: hairs) from rete roots. Therefore, for some lesions (especially lentigo cases) the density of retes will be inflated and consequently the proportion of the length of sections with loss of rete ridges to the total lesion length will be understated.
- In some cases the morphology of lentigo is similar to the morphology of melanoma: in both lesions one may observe the loss of rete ridges. They could still be distinguished one from another, as in melanoma typically nests are observed (e.g., at the dermoepidermal junction) and there is a pagetoid

spread of melanocytes whereas in lentigo no nests are present and melanocytes are arranged in lentiginous pattern, however, the proposed classification algorithm does not analyze neither nests nor patterns of melanocytes (since it was not the goal of our study).

8. Summary

This chapter points out the original contribution of the Author, summarizes the main conclusions drawn from the study, and outlines future research directions.

8.1. Original contribution

The most important original contribution of this study is the working out, implementation and testing of a complete system for automated diagnostics of skin melanocytic lesions in H&E-stained histopathological images (taken using optical microscopes) based on the analysis of lesion's epidermal morphometry. The proposed system scored the accuracy of 74.7% and 86.7% for the "lentigo vs. junctional dysplastic nevus vs. melanoma" and "benign vs. malignant" classification tasks, respectively. To the best of my knowledge it is the first attempt to automatically determine epidermal morphometry of a lesion and to utilize that information for the diagnostic purpose.

In order to develop the diagnostic system I proposed and verified algorithms, which constitute its successive stages, including:

- A tissue segmentation method (described in Chapter 4) which utilizes both global and local information about color distribution thus providing fine-grained, reliable, and easy to implement automatic foreground selection in high-resolution images of H&E-stained specimens, tolerant to the slide magnification, illumination conditions, artifacts and noise. Particularly notable is the use of global statistical information about color distribution in the image extracted from a 3D histogram of pixel intensities in the CIELAB color space.
- An epidermis segmentation method (described in Chapter 5) which utilizes information about shape and distribution of slide background regions in an image as well as information about distribution and concentration of hematoxylin and eosin stains extracted from their joint histogram (using statistical and image processing methods). To the best of my knowledge it is the only method which almost entirely depends on domain-specific knowledge of morphometric and biochemical properties of skin tissue elements to segment the relevant histopathological structures in human skin.
- A retes segmentation algorithm (described in Chapter 6) which automatically determines the basic morphometric features of individual rete ridges (i.e., their location, base width, length, and height) in a segmented epidermis. It is based mainly on analyzing the curvature of the epidermal boundary

to identify endpoints of projection bases, and on analyzing the skeleton of the epidermal region in the graph representation, using graph algorithms and morphological operations, to identify rete bases and the location of retes along the epidermis main axis. Most of the proposed methodology is a general-purpose solution which can be applied to a wide range of problems in computer vision related to shape analysis.

The above-mentioned methods are refined versions of algorithms I proposed and published as either journal articles or articles in conference proceedings:

- **P. Kłeczek**, J. Jaworek-Korjakowska, M. Gorgon. *A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images*. *Comput. Med. Imaging Graph.*, 2020, vol. 79, 2022, Art. ID 101686, doi: 10.1016/j.compmedimag.2019.101686 [IF (2018) = 3.298]
- **P. Kłeczek**, G. Dyduch, J. Jaworek-Korjakowska, R. Tadeusiewicz. *Automated epidermis segmentation in histopathological images of human skin stained with hematoxylin and eosin*. *Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology*, 101400M (2017), doi:10.1117/12.2249018 [WoS]
- **P. Kłeczek**, M. Lech, G. Dyduch, J. Jaworek-Korjakowska, R. Tadeusiewicz. *Segmentation of black ink and melanin in skin histopathological images*. *Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology*, 105811A (2018), doi:10.1117/12.2292859 [WoS]
- **P. Kłeczek**, G. Dyduch, A. Graczyk-Jarzynka, J. Jaworek-Korjakowska. *A New Approach to Border Irregularity Assessment with Application in Skin Pathology*. *Appl. Sci. (Basel)*, 2019, 9(10), 2022, doi:10.3390/app9102022 [IF (2018) = 2.217]

In my opinion, the most important contributions of my research to the field of computer vision are:

- The proposal of a methodology for image segmentation (in Section 5.1) which emphasizes the usefulness of the „void analysis”, i.e., the throughout analysis of image regions considered background regions, particularly based on domain-specific knowledge of morphometric properties of objects in analyzed images.
- The proposal of a novel general-purpose methodology of shape analysis (in Section 6.1), based mainly on graph algorithms and morphological operations applied on object’s skeleton, which can be used to identify projections on objects’ contour and to measure their morphometry.
- The proposal of general-purpose morphometric features of objects (in Section 7.1), which can be used to describe the jaggedness of object’s contour.

8.2. Conclusions

The main research aim of the proposed project was to determine the tissue structure of human epidermis, both normal and pathologically changed, in histopathological images of H&E-stained skin specimens and to identify changes in epidermal morphometry caused by nevi and melanomas using automated methods of digital image processing and machine learning. In the whole dissertation I unwaveringly aimed at reaching that goal and proving the following two theses proposed in Section 1.1: (1) using the

proposed methods for image segmentation and image analysis it is possible to assess the degree of malignancy of a given skin lesion, and (2) by means of machine learning methods it is possible to formulate the basic histopathological criteria for diagnosing skin lesions.

In order to do so, I developed (worked out, implemented and tested) a complete system for automated diagnostics of skin melanocytic lesions in H&E-stained histopathological images (taken using optical microscopes) based on the analysis of lesion's epidermal morphometry. That system, which is the most significant practical accomplishment of my study, was designed to distinguish between the following three types of skin melanocytic lesions: lentiginos (benign lesions), dysplastic junctional nevi (continua between benign and malignant lesions) and melanomas (malignant lesions). It is composed of four modules responsible for carrying out the following successive stages of the automatic diagnostic workflow: tissue segmentation, epidermis segmentation, rete segmentation, and lesion classification. The first three modules use automated methods of digital image processing, whereas the last one uses machine learning. The algorithms of all four aforementioned modules are fully automated and to a large extent based on domain-specific knowledge of morphometric and biochemical properties of skin tissue structures.

The performance of the whole system was verified on a dataset consisting of 75 high-resolution WSI images of skin melanocytic lesions (25 for each lesion type) diagnosed by an expert dermatopathologist. It scored the accuracy of 74.7% and 86.7% for the "lentigo vs. junctional dysplastic nevus vs. melanoma" and "benign vs. malignant" classification tasks, respectively. Therefore, I imply that the proposed system is capable of assessing the metastatic potential of a given skin lesion. In Section 7.3 a classification tree model (constructed based on morphometric indexes I proposed) is presented, which scored the accuracy of 73.3% and 88.0% for the "lentigo vs. junctional dysplastic nevus vs. melanoma" and "benign vs. malignant" classification tasks, respectively. That model could be used to help formulate the basic histopathological criteria related to epidermal morphometry.

The proposed tissue segmentation algorithm performs the foreground/background separation accurately and consistently, regardless of resolution and staining quality of images, with an overall area overlap of more than 92.9% and a deviation of 5.3% in the overlap. The epidermis segmentation method scored the mean sensitivity 96.5%, specificity 96.5% and precision 69.5% (it also successfully yielded meaningful results in all tested cases). The proposed rete segmentation algorithm was verified for two sets of epidermis segmentation masks: the first consisted of manual segmentations prepared by an expert dermatopathologist, whereas the second consisted of segmentations obtained automatically using the method described in Chapter 5). For manual segmentations, 83.2% of manually segmented retes were detected correctly, 14.8% of manually segmented retes were not detected, whereas 4.7% of detected retes were false positives (the rest cases were debatable). The "strict" Jaccard similarity coefficient for the task of detecting rete ridges was 0.798 (the main sources of errors were misdetection of micro-retes and treating skin appendages as retes). The mean absolute value of relative errors was 20.7% for the width and 16.6% for the length. For automatic segmentations, 42.4% of manually segmented retes were detected correctly, 52.1% of manually segmented retes were not detected, whereas 33.4% of detected retes

were false positives (the rest cases were debatable). However, such mediocre performance for automatic segmentations did not significantly affect the results obtained in the classification stage.

Not only is my diagnostic system useful as the whole, but also each of its segmentation modules has a wide range of applications. In particular, the tissue segmentation method is usually an obligatory first step in computerized analysis of WSIs (not only in digital dermatopathology), the algorithm for epidermis segmentation could be used to increase effectiveness of other diagnostics algorithms (e.g., those detecting and analyzing distribution of melanocytes in the epidermis), and the rete segmentation technique could be used to analyze morphometry of other types of skin lesions (e.g., inflammations).

Moreover, I consider methods proposed in this dissertation an important contribution to the field of computer vision. They present a new methodology for image segmentation which emphasizes the usefulness of the „void analysis” (i.e., the throughout analysis of image regions considered background regions, particularly based on domain-specific knowledge of morphometric properties of objects in analyzed images), a novel general-purpose methodology of shape analysis (based mainly on graph algorithms and morphological operations applied on object’s skeleton) useful for the identification of projections on objects’ contour and for their morphometry measurement, as well as general-purpose morphometric features of objects (which can be used to describe the jaggedness of object’s contour).

In my opinion the obtained results let me pronounce that the main research aim of the proposed project was reached and that both theses were proven.

8.3. Future research directions

The proposed automatic diagnostic algorithm and each of its stages were tested on a large dataset of 60 H&E-stained WSIs images. Even though the obtained segmentation and classification results were satisfactory, all the methods should be verified on a larger dataset consisting not only more cases, but also images from various sources (currently only the tissue and epidermis segmentation methods were tested using images from various sources).

The epidermis segmentation algorithm should be refined so that it would deal better with abundant lymphohistiocytic infiltrates, regions of dense connective tissue (such as homogenized collagen), hyperkeratosis, and isolated skin appendages present in the dermis. Currently, the presence of these structures and/or pathological changes typically decreases the precision rate of the resulting segmentation. The retes segmentation algorithm should be polished up so that it would detect epidermal projections being in fact skin appendages, which are currently considered by the algorithm as retes (due to morphological similarity of both structures). To the best of my knowledge no such a skin appendages segmentation method was proposed so far.

Finally, the diagnostic algorithm could be extended with a routine to segment nests of melanocytes and integrated with existing algorithms for the detection of melanocytes (such as the one proposed by Xu et al. [133]), as the information about their presence and distribution pattern provide valuable diagnostic information.

Bibliography

- [1] V. Kumar, A. K. Abbas, and N. Fausto. *Robbins and Cotran Pathologic Basis of Disease*. Ed. by V. Kumar et al. 7th ed. Vol. 1. Elsevier Saunders, 2005. ISBN: 978-1-4160-3121-5.
- [2] S. Lodha et al. “Discordance in the histopathologic diagnosis of difficult melanocytic neoplasms in the clinical setting”. In: *Journal of Cutaneous Pathology* 35.4 (2008), pp. 349–352. DOI: [10.1111/j.1600-0560.2007.00970.x](https://doi.org/10.1111/j.1600-0560.2007.00970.x).
- [3] M. Lin et al. “Diagnostic accuracy of malignant melanoma according to subtype”. In: *Australasian Journal of Dermatology* 55.1 (2013), pp. 35–42.
- [4] D. Snead et al. “Validation of digital pathology imaging for primary histopathological diagnosis”. In: *Histopathology* 68.7 (2016), pp. 1063–1072. DOI: [10.1111/his.12879](https://doi.org/10.1111/his.12879).
- [5] C. Garbe and U. Leiter. “Melanoma epidemiology and trends”. In: *Clin Dermatol* 27.1 (2009), pp. 3–9. DOI: [10.1016/j.clindermatol.2008.09.001](https://doi.org/10.1016/j.clindermatol.2008.09.001).
- [6] *Cancer Facts & Figures 2016*. American Cancer Society. 2016. URL: <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index> (visited on 2020-01-10).
- [7] Australian Bureau of Statistics. *3303.0 Causes of death, Australia 2015*. 2016.
- [8] *Zachorowania i zgony na nowotwory złośliwe w Polsce*. Krajowy Rejestr Nowotworów, Centrum Onkologii - Instytut im. Marii Skłodowskiej-Curie. 2016. URL: <http://onkologia.org.pl/raporty/> (visited on 2020-01-10).
- [9] D. M. Lipsker et al. “Striking Increase of Thin Melanomas Contrasts With Stable Incidence of Thick Melanomas”. In: *Archives of Dermatology* 135.12 (1999), pp. 1451–1456. DOI: [10.1001/archderm.135.12.1451](https://doi.org/10.1001/archderm.135.12.1451). eprint: [/data/Journals/DERM/11693/DST9023.pdf](http://data.journals.dermatology.com/11693/DST9023.pdf).
- [10] *Skin cancer incidence trends over time*. Cancer Research UK. 2016. URL: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence%5C#heading-Two> (visited on 2020-01-10).
- [11] *SEER Stat Fact Sheets: Melanoma of the Skin*. National Cancer Institute. 2016. URL: <https://seer.cancer.gov/statfacts/html/melan.html> (visited on 2020-01-10).
- [12] G. Argenziano et al. *Interactive Atlas of Dermoscopy*. Milan, Italy: Edra Medical Publishing and New Media, 2000. ISBN: 978-8-88645-730-9.

- [13] P. H. Soyer et al. *Color atlas of melanocytic lesions of the skin*. Springer, 2007.
- [14] B. Shoo, R. Sagebiel, and M. Kashani-Sabet. “Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center”. In: *Journal of the American Academy of Dermatology* 62.5 (2010), pp. 751–756.
- [15] S. Patrawala et al. “Discordance of histopathologic parameters in cutaneous melanoma: Clinical implications”. In: *Journal of the American Academy of Dermatology* 74.1 (2016), pp. 75–80. ISSN: 0190-9622. DOI: <http://dx.doi.org/10.1016/j.jaad.2015.09.008>.
- [16] G. Massi and P. E. LeBoit. *Histological Diagnosis of Nevi and Melanoma*. Springer-Verlag Berlin Heidelberg, 2014. DOI: [10.1007/978-3-642-37311-4](https://doi.org/10.1007/978-3-642-37311-4).
- [17] R. L. Barnhill et al. “Cutaneous Melanoma”. In: *Pathology of Melanocytic Nevi and Melanoma*. Ed. by R. L. Barnhill, M. Piepkorn, and K. J. Busam. Springer-Verlag Berlin Heidelberg, 2014. Chap. 10, pp. 331–488.
- [18] M. Filho, Z. Ma, and J. M. Tavares. “A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices”. In: *Journal of Medical Systems* 39.11 (2015), pp. 1–12. ISSN: 0148-5598. DOI: [10.1007/s10916-015-0354-8](https://doi.org/10.1007/s10916-015-0354-8).
- [19] R. B. Oliveira et al. “Computational methods for the image segmentation of pigmented skin lesions: A review”. In: *Computer Methods and Programs in Biomedicine* 131 (2016), pp. 127–141. ISSN: 0169-2607. DOI: <https://doi.org/10.1016/j.cmpb.2016.03.032>.
- [20] R. B. Oliveira et al. “Computational Methods for Pigmented Skin Lesion Classification in Images: Review and Future Trends”. In: *Neural Computing and Applications* 29.3 (2018), pp. 613–636. ISSN: 0941-0643. DOI: [10.1007/s00521-016-2482-6](https://doi.org/10.1007/s00521-016-2482-6).
- [21] Y. Yuan, M. Chao, and Y. Lo. “Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance”. In: *IEEE Transactions on Medical Imaging* 36.9 (2017), pp. 1876–1886. ISSN: 0278-0062. DOI: [10.1109/TMI.2017.2695227](https://doi.org/10.1109/TMI.2017.2695227).
- [22] M. H. Jafari et al. “Skin lesion segmentation in clinical images using deep learning”. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. 2016, pp. 337–342. DOI: [10.1109/ICPR.2016.7899656](https://doi.org/10.1109/ICPR.2016.7899656).
- [23] Y. Li and L. Shen. “Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network”. In: *Sensors* 18.2 (2018). ISSN: 1424-8220. DOI: [10.3390/s18020556](https://doi.org/10.3390/s18020556).
- [24] P. Kłeczek et al. “Automated epidermis segmentation in histopathological images of human skin stained with hematoxylin and eosin”. In: *Proc. SPIE 10140, Medical Imaging 2017: Digital Pathology, 101400M*. 2017. DOI: [doi:10.1117/12.2249018](https://doi.org/10.1117/12.2249018).
- [25] P. Kłeczek, J. Jaworek-Korjakowska, and M. Gorgon. “A novel method for tissue segmentation in high-resolution H&E-stained histopathological whole-slide images”. In: *Computerized Medical Imaging and Graphics* 79 (2020), pp. 1–12. ISSN: 0895-6111. DOI: <https://doi.org/10.1016/j.compmedimag.2019.101686>.

- [26] P. Kleczek et al. “A New Approach to Border Irregularity Assessment with Application in Skin Pathology”. In: *Applied Sciences* 9.10 (2019). ISSN: 2076-3417. DOI: [10.3390/app9102022](https://doi.org/10.3390/app9102022).
- [27] P. Kleczek et al. “Segmentation of black ink and melanin in skin histopathological images”. In: *Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 105811A*. 2018. DOI: <https://doi.org/10.1117/12.2292859>.
- [28] W. Pawlina and M. H. Ross. *Histology: A Text and Atlas*. 7th ed. LWW, 2016.
- [29] D. S. Strayer and E. Rubin. “Cell Adaptation, Cell Injury and Cell Death”. In: *Rubin’s Pathology: Clinicopathologic Foundations of Medicine*. Ed. by R. Rubin, D. S. Strayer, and E. Rubin. 6th ed. Baltimore, Philadelphia: Lippincott Williams & Wilkins, 2011. Chap. 1, pp. 1–46.
- [30] R. P. Braun et al. “Agreement of dermatopathologists in the evaluation of clinically difficult melanocytic lesions: how golden is the ‘gold standard’?” In: *Dermatology* 224.1 (2012), pp. 51–58. ISSN: 1018-8665.
- [31] M. E. Celebi, N. Codella, and A. Halpern. “Dermoscopy Image Analysis: Overview and Future Directions”. In: *IEEE Journal of Biomedical and Health Informatics* 23.2 (2019), pp. 474–478. DOI: [10.1109/JBHI.2019.2895803](https://doi.org/10.1109/JBHI.2019.2895803).
- [32] G. Rolls. *An Introduction to Specimen Preparation*. Leica Biosystems. 2011. URL: <https://www.leicabiosystems.com/pathologyleaders/an-introduction-to-specimen-preparation/> (visited on 2020-01-10).
- [33] J. D. Bancroft and C. Layton. “The hematoxylin and eosin”. In: *Bancroft’s Theory and Practice of Histological Techniques*. Ed. by S. K. Suvarna, C. Layton, and J. D. Bancroft. 7th ed. Churchill Livingstone, 2013. Chap. 10, pp. 173–186.
- [34] J. K. Chan. “The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology”. In: *International Journal of Surgical Pathology* 22.1 (2014). ISSN: 1940-2465. DOI: [10.1177/1066896913517939](https://doi.org/10.1177/1066896913517939).
- [35] S. Chatterjee. “Artefacts in histopathology”. In: *Journal of Oral and Maxillofacial Pathology* 18.Suppl 1 (2014), S111–S116. DOI: [10.4103/0973-029X.141346](https://doi.org/10.4103/0973-029X.141346).
- [36] S. Taqi et al. “A review of artifacts in histopathology”. In: *Journal of Oral and Maxillofacial Pathology* 22.2 (2018), p. 279. DOI: [10.4103/jomfp.JOMFP_125_15](https://doi.org/10.4103/jomfp.JOMFP_125_15).
- [37] M. Srinivasan, D. Sedmak, and S. Jewell. “Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids”. In: *The American Journal of Pathology* 161.6 (2002), pp. 1961–1971. DOI: <http://dx.doi.org/10.1118/1.4939130>.
- [38] J. E. Margarone, J. R. Natiella, and C. D. Vaughan. “Artifacts in oral biopsy specimens”. In: *Journal of Oral and Maxillofacial Surgery* 43.3 (1985), pp. 163–172. DOI: <http://dx.doi.org/10.1118/1.4939130>.

- [39] P. Krishnanand et al. “Artefacts in Oral Mucosal Biopsies: A Review”. In: *Journal of Orofacial Sciences* 2.1 (2010), pp. 57–62.
- [40] V. Rastogi et al. “Artefacts: a diagnostic dilemma - a review”. In: *Journal of Clinical and Diagnostic Research* 7.10 (2010), pp. 2408–2413. DOI: [10.7860/JCDR/2013/6170.3541](https://doi.org/10.7860/JCDR/2013/6170.3541).
- [41] L. Pantanowitz et al. “Review of the current state of whole slide imaging in pathology”. In: *Journal of Pathology Informatics* 2.1 (2011), p. 36. DOI: [10.4103/2153-3539.83746](https://doi.org/10.4103/2153-3539.83746).
- [42] P. S. Nielsen et al. “Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors”. In: *Human Pathology* 41.12 (2010), pp. 1770–1776. ISSN: 0046-8177. DOI: [10.1016/j.humpath.2010.05.015](https://doi.org/10.1016/j.humpath.2010.05.015).
- [43] ZEISS Axio Scan.Z1. ZEISS International. 2019. URL: <https://www.zeiss.com/microscopy/us/products/imaging-systems/axio-scan-z1.html> (visited on 2019-11-12).
- [44] J. Ho et al. “Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization”. In: *Journal of Pathology Informatics* 5.33 (2014), pp. 1–14. DOI: [10.4103/2153-3539.139714](https://doi.org/10.4103/2153-3539.139714).
- [45] M. D. Feldman. “Beyond Morphology: Whole Slide Imaging, Computer-Aided Detection, and Other Techniques”. In: *Archives of Pathology & Laboratory Medicine* 132.5 (2008), pp. 758–763. DOI: [10.1043/1543-2165\(2008\)132\[758:BMWSIC\]2.0.CO;2](https://doi.org/10.1043/1543-2165(2008)132[758:BMWSIC]2.0.CO;2). eprint: <https://www.archivesofpathology.org/doi/pdf/10.1043/1543-2165%282008%29132%5B758%3ABMWSIC%5D2.0.CO%3B2>.
- [46] R. C. Friedberg and L. Pantanowitz. “Practice Evolution: Decentralized Computer-Assisted Immunohistochemical Image Analysis”. In: *Archives of Pathology & Laboratory Medicine* 133.4 (2009), pp. 597–600. DOI: [10.1043/1543-2165-133.4.597](https://doi.org/10.1043/1543-2165-133.4.597). eprint: <https://www.archivesofpathology.org/doi/pdf/10.1043/1543-2165-133.4.597>.
- [47] A. Parwani et al. “Regulatory barriers surrounding the use of whole slide imaging in the United States of America”. In: *Journal of Pathology Informatics* 5.1 (2014), pp. 1–38. DOI: [10.4103/2153-3539.143325](https://doi.org/10.4103/2153-3539.143325).
- [48] N. Farahani, A. V. Parwani, and L. Pantanowitz. “Whole slide imaging in pathology: advantages, limitations, and emerging perspectives”. In: *Pathology and Laboratory Medicine International* 2015.7 (2015), pp. 23–33. DOI: [10.2147/PLMI.S59826](https://doi.org/10.2147/PLMI.S59826).
- [49] S. Thorstenson, J. Molin, and C. Lundström. “Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013”. In: *Journal of Pathology Informatics* 5.1 (2014), p. 14. DOI: [10.4103/2153-3539.129452](https://doi.org/10.4103/2153-3539.129452).
- [50] N. Stathonikos et al. “Going fully digital: Perspective of a Dutch academic pathology lab”. In: *Journal of Pathology Informatics* 4.1 (2013), p. 15. DOI: [10.4103/2153-3539.114206](https://doi.org/10.4103/2153-3539.114206).

- [51] M. Cree and H. Jelinek. “The effect of JPEG compression on automated detection of microaneurysms in retinal images”. In: *Proc. SPIE 6813, Image Processing: Machine Vision Applications, 68130M*. Vol. 6813. 2008. DOI: [10.1117/12.766276](https://doi.org/10.1117/12.766276).
- [52] *AXIO Imager.M2m*. LabX.com. 2019. URL: <https://www.labx.com/item/zeiss-axio-imager-m2m-motorized-microscope-dic-polarization/11071907> (visited on 2019-11-12).
- [53] F. Ghaznavi et al. “Digital Imaging in Pathology: Whole-Slide Imaging and Beyond”. In: *Annual Review of Pathology: Mechanisms of Disease*. Ed. by A. Abbas, S. Galli, and P. Howley. Vol. 8. Annual Review of Pathology-Mechanisms of Disease. 4139 El Camino Way, PO BOX 10139, Palo Alto, CA 94303-0897 USA: Annual Reviews, 2013, pp. 331–359. ISBN: 978-0-8243-4308-8. DOI: [10.1146/annurev-pathol-011811-120902](https://doi.org/10.1146/annurev-pathol-011811-120902).
- [54] D. M. Jukic et al. “Clinical Examination and Validation of Primary Diagnosis in Anatomic Pathology Using Whole Slide Digital Images”. English. In: *Archives of Pathology & Laboratory Medicine* 135.3 (2011), pp. 372–378.
- [55] J. R. Gilbertson et al. “Primary histologic diagnosis using automated whole slide imaging: a validation study”. In: *BMC Clinical Pathology* 6.1 (2006), pp. 1–19. ISSN: 1472-6890. DOI: [10.1186/1472-6890-6-4](https://doi.org/10.1186/1472-6890-6-4).
- [56] S. Al-Janabi et al. “Whole slide images for primary diagnostics in dermatopathology: a feasibility study”. In: *Journal of Clinical Pathology* (2011). DOI: [10.1136/jclinpath-2011-200277](https://doi.org/10.1136/jclinpath-2011-200277).
- [57] N. Alsubaie et al. “Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation”. In: *PLOS ONE* 12.1 (Jan. 2017), pp. 1–15. DOI: [10.1371/journal.pone.0169875](https://doi.org/10.1371/journal.pone.0169875).
- [58] M. Macenko et al. “A method for normalizing histology slides for quantitative analysis”. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2009, pp. 1107–1110. DOI: [10.1109/ISBI.2009.5193250](https://doi.org/10.1109/ISBI.2009.5193250).
- [59] E. Reinhard et al. “Color transfer between images”. In: *IEEE Computer Graphics and Applications* 21.5 (2001), pp. 34–41. ISSN: 0272-1716. DOI: [10.1109/38.946629](https://doi.org/10.1109/38.946629).
- [60] M. Niethammer et al. “Appearance Normalization of Histology Slides”. In: *Machine Learning in Medical Imaging: First International Workshop, MLMI 2010, Held in Conjunction with MICCAI 2010, Beijing, China, September 20, 2010. Proceedings*. Ed. by F. Wang et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 58–66. ISBN: 978-3-642-15948-0. DOI: [10.1007/978-3-642-15948-0_8](https://doi.org/10.1007/978-3-642-15948-0_8).
- [61] A. C. Ruifrok and D. A. Johnston. “Quantification of histochemical staining by color deconvolution”. In: *Analytical and Quantitative Cytology and Histology* 23.4 (2001), pp. 291–299.
- [62] M. Gavrilovic et al. “Blind Color Decomposition of Histological Images”. In: *IEEE Transactions on Medical Imaging* 32.6 (2013), pp. 983–994. ISSN: 0278-0062. DOI: [10.1109/TMI.2013.2239655](https://doi.org/10.1109/TMI.2013.2239655).

- [63] A. M. Khan et al. “A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution”. In: *IEEE Transactions on Biomedical Engineering* 61.6 (2014), pp. 1729–1738. ISSN: 0018-9294. DOI: [10.1109/TBME.2014.2303294](https://doi.org/10.1109/TBME.2014.2303294).
- [64] N. Alsubaie et al. “A Discriminative Framework for Stain Deconvolution of Histopathology Images in the Maxwellian Space”. In: *Medical Image Understanding and Analysis - MIUA 2015. 19th Annual Conference, University of Lincoln, Lincoln, UK, July 15-17, 2015*. Ed. by T. Lambrou and X. Ye. BMVA, 2015, pp. 132–137.
- [65] J. N. Kather et al. “New Colors for Histology: Optimized Bivariate Color Maps Increase Perceptual Contrast in Histological Images”. In: *PLOS ONE* 10.12 (Dec. 2016), pp. 1–22. DOI: [10.1371/journal.pone.0145572](https://doi.org/10.1371/journal.pone.0145572).
- [66] N. Trahearn et al. “Multi-class stain separation using independent component analysis”. In: *Proc. SPIE 9420, Medical Imaging 2015: Digital Pathology, 94200J*. Vol. 9420. 2015. DOI: [10.1117/12.2081933](https://doi.org/10.1117/12.2081933).
- [67] M. Veta et al. “Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge”. In: *Medical Image Analysis* 54 (2019), pp. 111–121. ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2019.02.012>.
- [68] M. Saraswat and K. V. Arya. “Automated microscopic image analysis for leukocytes identification: A survey”. In: *Micron* 65 (2014), pp. 20–33. ISSN: 0968-4328. DOI: <http://dx.doi.org/10.1016/j.micron.2014.04.001>.
- [69] D. Magee et al. “Colour Normalisation in Digital Histopathology Images”. In: *Proc. Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*. 2009, pp. 100–111.
- [70] H. Shimizu. “Structure and Function of the Skin”. In: *Shimizu’s Textbook of Dermatology*. 1st ed. Hokkaido University Press/Nakayama Shoten, 2007. Chap. 1, pp. 1–26.
- [71] C. G. Lian and G. F. Murphy. “Histology of the Skin”. In: *Lever’s Histopathology of the Skin*. LWW, 2015, pp. 8–75. ISBN: 978-1451190373.
- [72] T. Igarashi, K. Nishino, and S. K. Nayar. “The Appearance of Human Skin: A Survey”. In: *Foundations and Trends in Computer Graphics and Vision* 3.1 (2007), pp. 1–95. ISSN: 1572-2740. DOI: [10.1561/06000000013](https://doi.org/10.1561/06000000013).
- [73] P. Agache and P. Humbert. *Measuring the Skin*. Springer-Verlag Berlin Heidelberg, 2004. ISBN: 978-3-540-01771-4.
- [74] F. M. Watt and H. Green. “Involucrin Synthesis Is Correlated with Cell Size in Human Epidermal Cultures”. In: *The Journal of Cell Biology* 90 (3 1981), pp. 738–742.

- [75] D. Gareau. “Automated identification of epidermal keratinocytes in reflectance confocal microscopy”. In: *Journal of Biomedical Optics* 16 (3 2011). DOI: <https://dx.doi.org/10.1117%2F1.3552639>.
- [76] M. Huzaira et al. “Topographic Variations in Normal Skin, as Viewed by In Vivo Reflectance Confocal Microscopy”. In: *Journal of Investigative Dermatology* 116 (6 2001), pp. 846–852.
- [77] J. Sandby-Möller, T. Poulsen, and H. C. Wulf. “Epidermal Thickness at Different Body Sites: Relationship to Age, Gender, Pigmentation, Blood Content, Skin Type and Smoking Habits”. In: *Acta Dermato-Venereologica* 83 (2003), pp. 410–413.
- [78] A. Laurent et al. “Echographic measurement of skin thickness in adults by high frequency ultrasound to assess the appropriate microneedle length for intradermal delivery of vaccines”. In: *Vaccine* 25 (34 2007), pp. 6423–6430.
- [79] K. Robertson and J. L. Rees. “Variation in Epidermal Morphology in Human Skin at Different Body Sites as Measured by Reflectance Confocal Microscopy”. In: *Acta Dermato-Venereologica* 90 (2010), pp. 368–373.
- [80] J. K. Barton et al. “Investigating Sun-damaged Skin and Actinic Keratosis with Optical Coherence Tomography: A Pilot Study”. In: *Technology in Cancer Research & Treatment* 2.6 (2003), pp. 525–535.
- [81] C. Fournier et al. “Relationship between skin-collagen micro-architecture and ultrasonic backscatter parameters at 20 MHz”. In: *Ultrasonics Symposium, 2002. Proceedings. 2002 IEEE*. Vol. 1. 2002, pp. 625–628. DOI: [10.1109/ULTSYM.2002.1193480](https://doi.org/10.1109/ULTSYM.2002.1193480).
- [82] M. W. Piepkorn and R. L. Barnhill. “Common Acquired and Atypical/Dysplastic Melanocytic Nevi”. In: *Pathology of Melanocytic Nevi and Melanoma*. Ed. by R. L. Barnhill, M. Piepkorn, and K. J. Busam. Springer-Verlag Berlin Heidelberg, 2014. Chap. 5, pp. 87–154.
- [83] R. M. MacKie. “Disorders of the Cutaneous Melanocyte”. In: *Rook’s Textbook of Dermatology*. Ed. by T. Burns et al. 7th ed. Blackwell Publishing, 2004. Chap. 38.
- [84] S. F. Cramer. “The origin of epidermal melanocytes. Implications for the histogenesis of nevi and melanomas”. In: *Archives of Pathology and Laboratory Medicine* 115.2 (1991).
- [85] J. M. Grichnik. “Melanoma, Nevogenesis, and Stem Cell Biology”. In: *Journal of Investigative Dermatology* 128.10 (2008), pp. 2365–2380. ISSN: 0022-202X. DOI: <https://doi.org/10.1038/jid.2008.166>.
- [86] M. C. Mihm and M. J. Imber. “Benign melanocytic tumors”. In: *Pathology of the skin*. Ed. by E. R. Farmer and A. F. Hood. Appleton & Lange, 1990, pp. 663–683. ISBN: 9780838577158.
- [87] *Mosby’s Medical Dictionary*. 10th ed. Elsevier, 2016. ISBN: 9780323414258.

- [88] A. C. Allen. "A reorientation on the histogenesis and clinical significance of cutaneous nevi and melanomas". In: *Cancer* 2.1 (1949), pp. 28–56. DOI: [10.1002/1097-0142\(194901\)2:1<28::AID-CNCR2820020105>3.0.CO;2-N](https://doi.org/10.1002/1097-0142(194901)2:1<28::AID-CNCR2820020105>3.0.CO;2-N). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142%28194901%292%3A1%3C28%3A%3AAID-CNCR2820020105%3E3.0.CO%3B2-N>.
- [89] P. Masson. "My conception of cellular nevi". In: *Cancer* 4.1 (1951), pp. 9–38. DOI: [10.1002/1097-0142\(195101\)4:1<9::AID-CNCR2820040104>3.0.CO;2-0](https://doi.org/10.1002/1097-0142(195101)4:1<9::AID-CNCR2820040104>3.0.CO;2-0). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142%28195101%294%3A1%3C9%3A%3AAID-CNCR2820040104%3E3.0.CO%3B2-0>.
- [90] A. Shors et al. "Dysplastic naevi with moderate to severe histological dysplasia: a risk factor for melanoma". In: *British Journal of Dermatology* 155.5 (2006), pp. 988–993. DOI: [10.1111/j.1365-2133.2006.07466.x](https://doi.org/10.1111/j.1365-2133.2006.07466.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2133.2006.07466.x>.
- [91] L. Naldi et al. "Sun Exposure, Phenotypic Characteristics, and Cutaneous Malignant Melanoma. An Analysis According to Different Clinico-Pathological Variants and Anatomic Locations (Italy)". In: *Cancer Causes & Control* 16.8 (2005), pp. 893–899. ISSN: 1573-7225. DOI: [10.1007/s10552-005-2300-4](https://doi.org/10.1007/s10552-005-2300-4).
- [92] P. J. Heenan, D. E. Elder, and L. H. Sobin. *Histological Typing of Skin Tumours*. 2nd ed. Springer, 1996.
- [93] J. Clark Wallace H. et al. "Origin of Familial Malignant Melanomas From Heritable Melanocytic Lesions: 'The B-K Mole Syndrome'". In: *JAMA Dermatology* 114.5 (1978), pp. 732–738. ISSN: 2168-6068. DOI: [10.1001/archderm.1978.01640170032006](https://doi.org/10.1001/archderm.1978.01640170032006). eprint: https://jamanetwork.com/journals/jamadermatology/articlepdf/538650/archderm_114_5_006.pdf.
- [94] "Consensus Conference on Precursors to Malignant Melanoma". In: *The Journal of Dermatologic Surgery and Oncology* 11.5 (1985), pp. 537–542. DOI: [10.1111/j.1524-4725.1985.tb01416.x](https://doi.org/10.1111/j.1524-4725.1985.tb01416.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1524-4725.1985.tb01416.x>.
- [95] M. Shapiro et al. "Variability in nomenclature used for nevi with architectural disorder and cytologic atypia (microscopically dysplastic nevi) by dermatologists and dermatopathologists". In: *Journal of Cutaneous Pathology* 31.8 (2004), pp. 523–530. DOI: [10.1111/j.0303-6987.2004.00216.x](https://doi.org/10.1111/j.0303-6987.2004.00216.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0303-6987.2004.00216.x>.
- [96] K. Korotkov and R. Garcia. "Computerized analysis of pigmented skin lesions: a review". In: *Artificial intelligence in medicine* 56.2 (2012), pp. 69–90.
- [97] N. M. Prince, A. M. Rywlin, and A. B. Ackerman. "Histologic criteria for the diagnosis of superficial spreading malignant melanoma: formulated on the basis of proven metastatic lesions". In: *Cancer* 38.6 (1976), pp. 2434–2441.

- [98] C. Urso et al. "Sensitivity and specificity of histological criteria in the diagnosis of conventional cutaneous melanoma". In: *Melanoma Research* 18.4 (2008), pp. 253–258. DOI: [10.1097/CMR.0b013e3283043cc0](https://doi.org/10.1097/CMR.0b013e3283043cc0).
- [99] S. W. Menzies et al. "The performance of SolarScan: an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma". In: *Archives of Dermatology* 141.11 (2005), pp. 1388–1396.
- [100] A. Masood and A. Ali Al-Jumaily. "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms". In: *International Journal of Biomedical Imaging* 2013 (2013), pp. 1–22.
- [101] E. R. Farmer, R. Gonin, and M. P. Hanna. "Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologist". In: *Human Pathology* 27.6 (1996), pp. 528–531.
- [102] D. Troxel. "An Insurer's Perspective on Error and Loss in Pathology". In: *Archives of Pathology & Laboratory Medicine* 129.10 (2005), pp. 1234–1236.
- [103] C. F. Heal et al. "Agreement between histological diagnosis of skin lesions by histopathologists and a dermato-histopathologist". In: *International Journal of Dermatology* 48.12 (2009), pp. 1366–1369.
- [104] S. W. McCarthy and R. A. Scolyer. "Pitfalls and Important Issues in the Pathologic Diagnosis of Melanocytic Tumors". In: *Ochsner Journal* 10.2 (2010), pp. 66–74.
- [105] D. Troxel. "Pitfalls in the diagnosis of malignant melanoma: findings of a risk management panel study". In: *American Journal of Surgical Pathology* 27.9 (2003), pp. 1278–1283.
- [106] D. Bug, F. Feuerhake, and D. Merhof. "Foreground Extraction for Histopathological Whole Slide Imaging". In: *Bildverarbeitung für die Medizin*. Ed. by H. Handels et al. Informatik Aktuell. Springer, 2015, pp. 419–424. ISBN: 978-3-662-46224-9.
- [107] P. Bándi et al. "Comparison of Different Methods for Tissue Segmentation in Histopathological Whole-Slide Images". In: *Computing Research Repository* abs/1703.05990 (2017). eprint: 1703.05990.
- [108] C. Rother, V. Kolmogorov, and A. Blake. "'GrabCut': Interactive Foreground Extraction Using Iterated Graph Cuts". In: *ACM Transactions on Graphics* 23.3 (2004), pp. 309–314. ISSN: 0730-0301. DOI: [10.1145/1015706.1015720](https://doi.org/10.1145/1015706.1015720).
- [109] C. A. Storm and D. E. Elder. "Skin". In: *Rubin's Pathology: Clinicopathologic Foundations of Medicine*. Ed. by R. Rubin, D. S. Strayer, and E. Rubin. 6th ed. Baltimore, Philadelphia: Lippincott Williams & Wilkins, 2011. Chap. 24, pp. 1111–1168.
- [110] M. Mokhtari et al. "Computer aided measurement of melanoma depth of invasion in microscopic images". In: *Micron* 61 (2014), pp. 40–48. ISSN: 0968-4328. DOI: <https://doi.org/10.1016/j.micron.2014.02.001>.

- [111] J. Haggerty et al. “Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin”. In: *BMC Medical Imaging* 14.1 (2014), pp. 1–21. DOI: [10.1186/1471-2342-14-7](https://doi.org/10.1186/1471-2342-14-7).
- [112] C. Lu and M. Mandal. “Automated segmentation and analysis of the epidermis area in skin histopathological images”. In: *Conf Proc IEEE Eng Med Biol Soc.* 2012, pp. 5355–5359. DOI: [10.1109/EMBC.2012.6347204](https://doi.org/10.1109/EMBC.2012.6347204).
- [113] H. Xu and M. Mandal. “Epidermis segmentation in skin histopathological images based on thickness measurement and k-means algorithm”. In: *EURASIP Journal on Image and Video Processing* 2015.1 (2015), pp. 1–14. DOI: [10.1186/s13640-015-0076-3](https://doi.org/10.1186/s13640-015-0076-3).
- [114] L. Ogiela, R. Tadeusiewicz, and M. R. Ogiela. “Cognitive Analysis in Diagnostic DSS-Type IT Systems”. In: *Artificial Intelligence and Soft Computing – ICAISC 2006*. Ed. by L. Rutkowski et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 962–971. ISBN: 978-3-540-35750-6.
- [115] R. Tadeusiewicz, L. Ogiela, and M. R. Ogiela. “Cognitive Analysis Techniques in Business Planning and Decision Support Systems”. In: *Artificial Intelligence and Soft Computing – ICAISC 2006*. Ed. by L. Rutkowski et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1027–1039. ISBN: 978-3-540-35750-6.
- [116] *University of Michigan Virtual Slide Box*. <https://www.pathology.med.umich.edu/slides/search.php?collection=Andea&dxview=show>. University of Michigan, Department of Pathology, 2013.
- [117] *UBC Virtual Slidebox*. <http://histo.anat.ubc.ca/PATHOLOGY/Anatomical%20Pathology/DermPath/>. University of British Columbia, Faculty of Medicine, 2005.
- [118] D. Alman et al. *CIE Publication 142: Improvement to industrial colour-difference evaluation*. Tech. rep. CIE, 2001.
- [119] G. Sharma, W. Wu, and E. N. Dalal. “The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations”. In: *Color Research & Application* 30.1 (2005), pp. 21–30. DOI: [10.1002/col.20070](https://doi.org/10.1002/col.20070). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/col.20070>.
- [120] D. Alman et al. *CIE Publication 116: Industrial colour-difference evaluation*. Tech. rep. CIE, 1995.
- [121] M. Mahy, L. van Eycken, and A. Oosterlinck. “Evaluation of Uniform Color Spaces Developed after the Adoption of CIELAB and CIELUV”. In: *Color Research & Application* 19 (2 1994), pp. 105–121. DOI: [10.1111/j.1520-6378.1994.tb00070.x](https://doi.org/10.1111/j.1520-6378.1994.tb00070.x).
- [122] A. B. Ackerman et al. *Histologic Diagnosis of Inflammatory Skin Diseases*. Williams & Wilkins, 2005.
- [123] D. Garrod and M. Chidgey. “Desmosome structure, composition and function”. In: *Biochimica et Biophysica Acta* 1778 (3 2008), pp. 572–587.

- [124] J. D. Bancroft and C. Layton. “Connective and mesenchymal tissues with their stains”. In: *Bancroft’s Theory and Practice of Histological Techniques*. Ed. by S. K. Suvarna, C. Layton, and J. D. Bancroft. 7th ed. Churchill Livingstone, 2013. Chap. 11, pp. 187–214.
- [125] J. A. McGrath, R. A. J. Eady, and F. M. Pope. “Anatomy and Organization of Human Skin”. In: *Rook’s Textbook of Dermatology*. Ed. by T. Burns et al. 7th ed. Blackwell Publishing, 2004. Chap. 3.
- [126] M. Veta et al. “Assessment of algorithms for mitosis detection in breast cancer histopathology images”. In: *Medical Image Analysis* 20 (1 2015), pp. 237–248.
- [127] E. W. Dijkstra. “A Note on Two Problems in Connexion with Graphs”. In: *Numerische Mathematik* 1.1 (1959), pp. 269–271. ISSN: 0029-599X.
- [128] A. Głowacz and Z. Głowacz. “Recognition of images of finger skin with application of histogram, image filtration and K-NN classifier”. In: *Biocybernetics and Biomedical Engineering* 36 (1 2016), pp. 95–101. DOI: <http://dx.doi.org/10.1016/j.bbe.2015.12.005>.
- [129] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB*. New Jersey: Pearson Prentice Hall, 2004.
- [130] A. Telea and J. J. van Wijk. “An augmented Fast Marching Method for computing skeletons and centerlines”. In: *VISSYM ’02 Proceedings of the symposium on Data Visualisation 2002* (Barcelona, Spain). 2002, pp. 251–260.
- [131] J. E. Bresenham. “Algorithm for computer control of a digital plotter”. In: *IBM Systems Journal* 4.1 (1965), pp. 25–30. ISSN: 0018-8670.
- [132] J. Jaworek-Korjakowska and P. Kłeczek. “Automatic Classification of Specific Melanocytic Lesions Using Artificial Intelligence”. In: *BioMed Research International* 2016 (2016), pp. 1–17.
- [133] H. Xu et al. “Automated analysis and classification of melanocytic tumor on skin whole slide images”. In: *Computerized Medical Imaging and Graphics* 66 (2018), pp. 124–134.
- [134] C. Lu et al. “Automated Segmentation of the Melanocytes in Skin Histopathological Images”. In: *IEEE Journal of Biomedical and Health Informatics* 17.2 (2013), pp. 284–296. ISSN: 2168-2194. DOI: [10.1109/TITB.2012.2199595](https://doi.org/10.1109/TITB.2012.2199595).
- [135] N. Noroozi and A. Zakerolhosseini. “Differential diagnosis of squamous cell carcinoma in situ using skin histopathological images”. In: *Computers in Biology and Medicine* 70 (2016), pp. 23–39. ISSN: 0010-4825. DOI: <https://doi.org/10.1016/j.combiomed.2015.12.024>.
- [136] T. G. Olsen et al. “Diagnostic performance of deep learning algorithms applied to three common diagnoses in dermatopathology”. In: *Journal of Pathology Informatics* 9.1 (2018), p. 32. DOI: [10.4103/jpi.jpi_31_18](https://doi.org/10.4103/jpi.jpi_31_18).
- [137] R. Tadeusiewicz. “Neural networks as a tool for modeling of biological systems”. In: *Bio-Algorithms and Med-Systems* 11.3 (2015), pp. 135–144. DOI: [10.1515/bams-2015-0021](https://doi.org/10.1515/bams-2015-0021).

- [138] J. Szalaniec et al. “Artificial neural network modelling of the results of tympanoplasty in chronic suppurative otitis media patients”. In: *Computers in Biology and Medicine* 43.1 (2013), pp. 16–22. ISSN: 0010-4825. DOI: <https://doi.org/10.1016/j.compbimed.2012.10.003>.
- [139] C. M. Bishop. *Pattern Recognition and Machine Learning*. Ed. by M. Jordan, J. Kleinberg, and B. Schölkopf. Information Science and Statistics. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0-387-31073-8.
- [140] M. Schmidt, N. Le Roux, and F. Bach. “Minimizing Finite Sums with the Stochastic Average Gradient”. In: *Mathematical Programming* 162.1-2 (2017), pp. 83–112. ISSN: 0025-5610. DOI: [10.1007/s10107-016-1030-6](https://doi.org/10.1007/s10107-016-1030-6).
- [141] P. Flach. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. New York, NY, USA: Cambridge University Press, 2012. ISBN: 1107422221.
- [142] L. Breiman et al. *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.
- [143] G. James et al. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Inc., 2014. ISBN: 1461471370.
- [144] K.-B. Duan and S. S. Keerthi. “Which Is the Best Multiclass SVM Method? An Empirical Study”. In: *Multiple Classifier Systems*. Ed. by N. C. Oza et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 278–285. ISBN: 978-3-540-31578-0.
- [145] S. Knerr, L. Personnaz, and G. Dreyfus. “Single-layer learning revisited: a stepwise procedure for building and training a neural network”. In: *Neurocomputing*. Ed. by F.-F. Soulié and J. Héroult. Vol. 68. NATO ASI Series. Springer Berlin Heidelberg, 1990, pp. 41–50. ISBN: 978-3-642-76155-3. DOI: [10.1007/978-3-642-76153-9_5](https://doi.org/10.1007/978-3-642-76153-9_5).
- [146] T. Hastie and R. Tibshirani. “Classification by Pairwise Coupling”. In: *Annals of Statistics* 26.2 (1998), pp. 451–471. DOI: [10.1214/aos/1028144844](https://doi.org/10.1214/aos/1028144844).
- [147] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. “Large Margin DAGs for Multiclass Classification”. In: *Advances in Neural Information Processing Systems*. MIT Press, 2000, pp. 547–553.
- [148] M. E. Celebi et al. “A methodological approach to the classification of dermoscopy images”. In: *Computerized Medical Imaging and Graphics* 31.6 (2007), pp. 362–373. ISSN: 08956111.
- [149] R.-E. Fan, P.-H. Chen, and C.-J. Lin. “Working Set Selection Using Second Order Information for Training Support Vector Machines”. In: *The Journal of Machine Learning Research* 6 (2005), pp. 1889–1918. ISSN: 1532-4435.
- [150] Y. Zhang and Y. Yang. “Cross-validation for selecting a model selection procedure”. In: *Journal of Econometrics* 187.1 (2015), pp. 95–112. ISSN: 0304-4076. DOI: <https://doi.org/10.1016/j.jeconom.2015.02.006>.

- [151] R. Kohavi. “A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection”. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence – Volume 2. IJCAI’95*. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143. ISBN: 1-55860-363-8.

Oświadczam, świadomy odpowiedzialności karnej za poświadczanie nieprawdy, że niniejszą pracę dyplomową wykonałem osobiście i samodzielnie i nie korzystałem ze źródeł innych niż wymienione w pracy.