Automatic Phenotyping of Microscopic Skin Images

By

Saif Raoof Hussein

A Thesis

Submitted for the Degree of Doctor of Philosophy

School of Science

University of Buckingham

2019

Buckingham, United Kingdom

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas, work or words have been included, I have adequately cited and referenced the original sources.

Saif Raoof Hussein

ACKNOWLEDGEMENTS

ABSTRACT

The Mouse Genetics Project (MGP) is a large-scale mutant mice production and phenotyping initiative that builds on the success of the Human Genome Project (HGP) to discover the functionality of all genes and their role in human diseases. The MGP aims to produce over 20,000 mutant lines of mouse model to investigate and quantify the impact of gene knockouts on the various organs. Due to the great deal of overlap between human and mouse genomes, most the acquired knowledge can be translated into diagnostics, biomarker identifications and eventually treatment for complex genetic-based diseases.

The skin is by far the largest organ in the mammalian body, and a complex structure of multiple layers consisting of different distinguishable cells and objects. Dermatology specialists have long associated many diseases with changes in different skin layers such as changes to the number of nuclei in the dermis and epidermis layers, the orientation of hair follicles, curvature of the outer border of the epidermis, and many more.

The manual quantification and analysis of such features/objects for a high throughput phenotype research project such as the MGP is an error-prone, time consuming and very costly in terms of resources and staff training. Recent rapid advances in biomedical image processing/analysis as well as the emergence of a variety of machine learning tools provide an exciting motivation to developing effective and efficient automatic solutions. Automation is as challenging as the manual methods but the challenges are of different nature. Therefore, this thesis is devoted to investigate, develop and test a number of automatic algorithms to quantify the above mentioned features/objects in mouse skin layers and experimentally identify the genetic causes of changes to these parameters in relation to skin diseases.

Our investigations and solutions had to deal with a number of technical challenges such as staining errors that lead colour overlapping between neighbouring layers/components, damages to the outer layers during preparation of images, the difficulty of establishing the ground truth for the large volume of image dataset, significant overlapping of nuclei objects, misalignments of tissues with the large volume of dataset. The main contributions of the thesis include:

1. Proposing an adaptive system that combines colour deconvolution and fuzzy c-mean methods to segment the three main skin layers in H&E images, namely the epidermis, dermis and fat cell layers, after establishing the limitation of existing solutions to overcome the challenges highlighted above.
2. Developing automatic methods for segmenting and counting nuclei in the epidermis and dermis layer in mice skin and demonstrating the ability of the proposal in identifying overlapping nuclei and separating them. Furthermore, we automatically identify a list of candidate genes responsible for abnormal changes to the number of nuclei.
3. Introducing a simple method to align the epidermis outer border in all images and designing an easy geometric-based formula to quantify the orientation of hair follicles with respect to the aligned epidermis as an indicator of skin abnormality. This led to identifying a list of candidate genes responsible for abnormal changes to the orientation of hair follicles.
4. Defining a simple mathematical model of the vaguely defined epidermis curvature as an indicator of skin abnormality and proposing a reliable scheme to quantify it.
5. Providing empirical evidences of the success for each of the above schemes by comparing the automatically determined quantification outputs with the ground truth determined by domain expert biology researchers.
6. Demonstrating the high potential of non-invasive, unsupervised machine learning techniques in the successful isolation of potentially interesting (knockout genes) relevant to genetic causes of skin abnormalities. This would facilitate high-throughput analysis in cutaneous research, with potential applications for screening drugs.

ABBREVIATIONS

APM	Arrestor pili muscles
ARL	Angle of Regression Line
BA	Bland-Altman
BCC	Basal cell carcinoma
CI	Confidence interval
CT	Computed tomography
CIELAB, L*a*b*	L* luminosity, a*&b* chromaticity colour space
DEJ	Dermal-epidermal junction
DNA	Deoxyribonucleic acid
DAB	3, 3′-Diaminobenzidine
EAC	Evidence Accumulation Clustering
ECM	Extracellular matrix
FCM	Fuzzy c-mean cluster
GUI	graphical user interfaces
HGP	Human Genome Project
HF	Hair follicle
H&E	Haematoxylin & eosin
HSV	Hue, saturation, value colour space
IMPC	International Mouse Phenotyping Consortium
IRS	Inner root sheath
KO	Knockout animals
LB	Lower Bound
LOCI	Laboratory for Optical and Computational Instrumentation (LOCI)
MAE	Mean Absolute Error
MGP	Mouse Genome Project
MRI	Magnetic resonance imaging
ORS	Outer root sheath
OD	Optical density
PAS	Periodic acid–Schiff
PCP	Planar cell polarity
PH	potential of Hydrogen
RR	Reference range
RGB	Red, green, blue colour space
ROI	Region of interest
SC	Stratum corneum
SG	Sebaceous gland
SVM	Support vector machine classifier
SPV	Single-pass voting
TIFF	Tagged Image File Format
UB	Upper Bound
UV	Ultraviolet
VSG	Sebaceous gland vector
VHF	Hair follicle vector
Wt	Wild-type
WTSI	Wellcome Trust Sanger Institute

Table of Contents

Table of Figures

Chapter 1 : Introduction

1.1 Research Introduction

Biological image processing and analysis techniques have increasingly become an essential source of tools that support scientists in evaluating the effects of environmental exposure and assessing physiological/biological changes in living beings tissues/organs by examining associated images captured with a variety of signals within the electromagnetic spectrum. In healthcare systems, the use of MRI, Ultrasound, Mammograms, X-rays and Gamma-rays images using a specialized sensors and devices are becoming a common practice for diagnostics and treatments (Suganya *et al.*, 2018). The use of these tools is naturally applicable to the evaluation of the effect of treatments and drug efficiency in the context of the drug discovery (Shuaib *et al.*, 2018). Microscopic images have for long been used by life scientists in their research endeavours to acquire and develop basic understanding of biological processes including the development of tumour and other type of tissue/organ abnormality, to be fed into medical practices as well as training (Bhattacharyya *et al.*, 2019). Over the last few decades, rapid growth in healthcare demand led to the emergence of high throughput research activities that are characterized by gigantic amounts of data/images to be analysed effectively at a rate beyond the capabilities of current research workforce (Wallentin and Lindahl, 2019). The growth in the number of large-scale research centers throughout the world especially after the successful completion of Human Genome Project is only a testimony to this phenomenon (Panagiotara *et al.*, 2019). However, the development in image processing/analysis techniques remained lagging behind due to a variety of reason including shortage of researchers in the field and lack of interest in multi-disciplinary activities, in particular (Sklansky and Bisconte, 2013). However, the rapid advances in computer vision and machine learning is creating ample opportunities for developing automatic image analysis techniques to meet the growing demands of high throughput research initiatives (Esteva *et al.*, 2019).

This thesis is meant to contribute to the development of automated microscopic image analysis techniques/tools for deployment in large scale biomedical research efforts for determining genetic causes of skin tissue abnormalities. In particular, the research project for this thesis came into being as part of discussions between Buckingham computing and biology researchers and their colleagues at the Sanger Institute, Cambridge, on automating certain image processing/analysis high throughput tasks conducted at the Sanger institute within their Mouse Genetics Project (MGP). The MGP is a large-scale mutant mice production and phenotyping initiative designed to integrate established knowledge of the mouse genome and of disease models. The techniques of diagnosis, and reducing time and laboratory costs, especially when there is a high throughput data need to analysis and process, such as our used data set. Where the data set in this thesis was 7000 H&E microscopic image(Takahashi *et al.*, 2019), which divided into two types, control type and mutant type.

1.2 General Description of the Research Project

The focus of this research project, being designed to contribute to the collaborative work described above in relation to the MGP, is naturally on investigations leading to the development of computational effective schemes for automatically measuring progressive structural changes in mouse skin suspected of being caused by pathological mechanisms. Such schemes are to be used to annotate microscopic H&E images of mice skin in order to overcome the capabilities of manual methods. Manual methods are a challenging task and it used by many research studies for a variety of skin features quantification.

The skin of a mammalian is the largest of its organs, and although apparently simple, it is a highly-organised tissue that is made up of three layers: subcutaneous adipose layer as the deep layer (Driskell *et al.*, 2014), the dermis layer forming a connective tissue above the adipose layer, and the epidermis as the top layer (Yousef and Sharma, 2018).

The organisation and structure of mammalian skin is widely researched for many aspects of dermatology and treatments usually assessed through quantitative measurements of certain features and objects, several of which change/deteriorate quantitatively due to disease or inactivated genes. Although the overall skin structure is consistent throughout the body, its thickness and other skin features are different depending on function and location (Honari, 2017). Cutaneous integrity is known to degrade with chronological age, a process exacerbated by environmental stress with cosmetic and pathological implications (Sauermann *et al.*, 2002). Strategies to mitigate or restore damaged matrix are highly desirable, thus objective morphological criteria and histological analysis are required to assess skin integrity. Most of the changed biological features, such as disorders, hyperplasia, underlying metabolic, connective tissue and inflammation can be revealed by close examination of the skin, i.e. the skin is the appropriate tissue to study.

Among the features of general interest that are relevant to our research are the number of *nuclei* presents in the epidermis layer as well as in the dermis layers, *hair follicles* embedded in the outer band of the epidermis, and the spatial structure of the epidermis surface. In all cases, enumeration and quantification of these objects and their shapes are indicators of skin abnormalities.

The main focus of our research is on quantifying changes to the number of nuclei in the dermis and epidermis layers, the orientation of hair follicles, curvature of the outer border of the epidermis. Accordingly, the research work conducted through my PhD program was devoted to investigate, design, and develop reliably effective automatic image processing/analysis algorithms to quantify the above mentioned features/objects extracted from images of mouse skin tissue within their known layers. Furthermore, machine learning based experiments were conducted to identify the genetic causes of changes to these parameters in relation to skin diseases.

Automating the detection of the above mentioned objects and quantifying the features can only be done after pre-processing the images by segmenting the different skin layers and their sub-layers. For microscopic images conventional segmentations of natural images are adversely affected by many factors including inadequate fixed image magnification, potential staining errors, overlapped objects, smugness, and/or missing parts.

Moreover, proving the effectiveness of any proposed scheme rely greatly on knowing the ground truth on a sufficiently large and diverse set of H&E skin microscopic images. Consultation with domain experts is necessary but is not straight forward. Moreover, interpreting and mimicking the way the trained life-science researchers or technicians present added requirements. The fuzzy nature of the way the human make relevant decisions is only one obvious difficulty. More specifically, we need appropriately computable functions to quantify concepts like shape of a nuclei-like object or a hair follicle, or the curvature of the complex epidermis layer.

In the analysis of skin histology, nuclei are important factors. For example, nuclei shape and organisation in the epidermis and dermis can qualitatively discriminate different ages, and measuring skin integrity to discriminate disease and a healthy state is an important tool for investigators (Brüne and Schiefenhövel, 2019).

The difficulty of the task and the consuming time in the classical manual analysis of skin morphology are the main motivation to contribute automatic methods, to measure alteration in two of three main skin layers by quantifying changes in the nuclei of each of two layers, and in particular the number and shape in the dermis and epidermis layers.

Overall in this study, we aimed to identify features that characterized disease-related changes in the skin, using novel automated methods to enhance, segment and analyse images acquired from microscope platforms in mouse samples, in biopsies technique.

For such research, a number of metrics are desired, including nuclei distribution (particularly changes in nuclei in different layers, dermis and epidermis), nuclei counting (to assess knock out genes function), quantifying the hair follicle orientation (for example, in assessing skin diseases), and measuring the curvature of epidermis layer (to assess the disorder of epidermis layer).

The reported work in the rest of the thesis have been greatly incentivised to exploit the growing new technology advances by combining computational non-machine learning methods and unsupervised machine learning of image processing techniques. We shall demonstrate, the benefits of this approach in (1) achieving the main objectives of this research project by facilitating the specific needs of the MPG for high-throughput image analysis tools for the quantification of the above-mentioned components of skin images, and in identifying genes causing certain skin abnormalities using the sufficiently large MGP knockout generated dataset of H&E skin mouse images. Ultimately, these analysis methods will facilitate accurate measurement of novel skin-restorative modalities.

1.3 The Research challenges

Throughout our investigations, we encountered many technical challenges in developing fully automatic mice annotation solution from H&E microscopic images. The main challenges, that could not be tackled by the above mentioned existing solutions, can be summarized as follows:

1- H&E staining issue. However the H&E stain is consistent and accurate but, it has several issues in term of the image quality, where the stain variation is the main tissue of the H&E stains i.e. the differences in pH (potential of Hydrogen), the stain age, usage degree, thickness and temperature (Larson *et al.*, 2011; Zhu *et al.*, 2015). All those factors may affect directly on the properties of the stain, which lead to make the processing of the image is difficult to process or untreatable, such as the following reasons.

- Information loss. Overlapping in the colour intensity of the background and foreground of image. Which causes loss of important information in the H&E images, such as nuclei in epidermis layers, which is displays in Figure 1.1 (A).
- Incorrect segmentation. The stain problems cause overlap colour between the skin layers that cause overlapping of the skin layers, such as overlap the epidermis and dermis layers. This lead to segment two layers (epidermis and part of dermis layers) as one layer, which is wrong segmentation as deems in Figure1.1 (B).

2- Variability placement of mouse skin layers. In the datasets we are dealing with, images show mouse skin layer placed in different angles. To make a fully automatic segmentation and analysis easier, the layers of skin needs to be aligned first. This step is mandatory in some cases in this thesis (as a part of Chapter 6). For example, automatic epidermis alignment to quantify the orientation of the hair follicles, which associated with the epidermis layer. The alignment was important stage to make the images comparable, because it needs to compare the orientation based on the angles of the follicles.

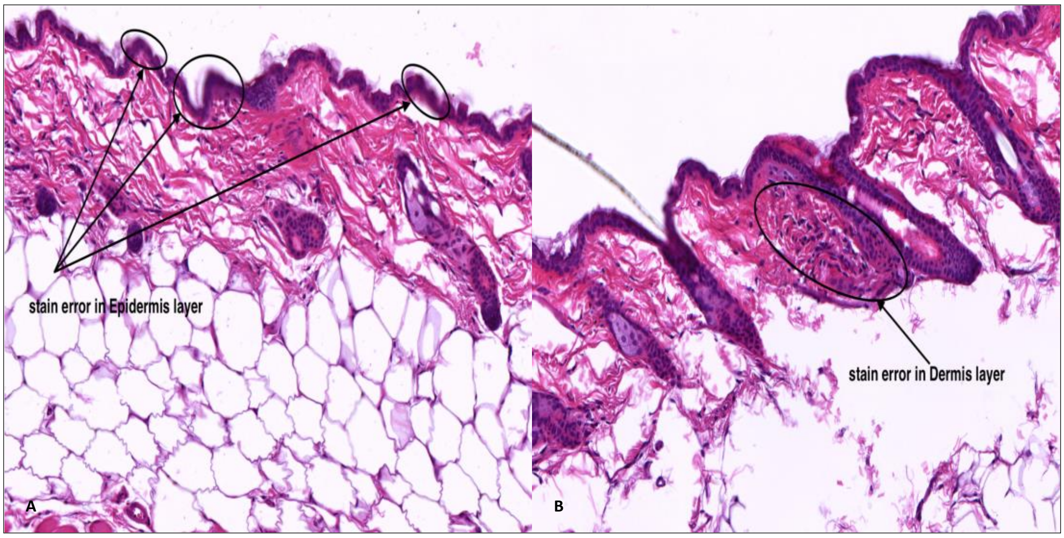


Figure 1.1 Stain error in different layers. A) Stain error in epidermis layer. B) Stain error in dermis layer

3. Ground truth. To evaluate our automatic methods, a robust ground truth that reports accurate measurements and associates diseases to images of the mouse is needed. In my thesis the ground truth I got for only in Chapter 6 the orientation of the hair follicle by expert annotators from Sanger Institute. In the other Chapters l

haven't got any ground truth from Sanger institute, such as ground truth for the nuclei numbers in different layers.

4. Initial segmentation of different ROIs. This research deals with different mouse layers. For example, Figure 1.2 (A, B) shows three different ROIs in three different mouse layer in the same view. Accurate annotations of the objects of interest rely on accurate and fully automatic initial segmentation of ROI.

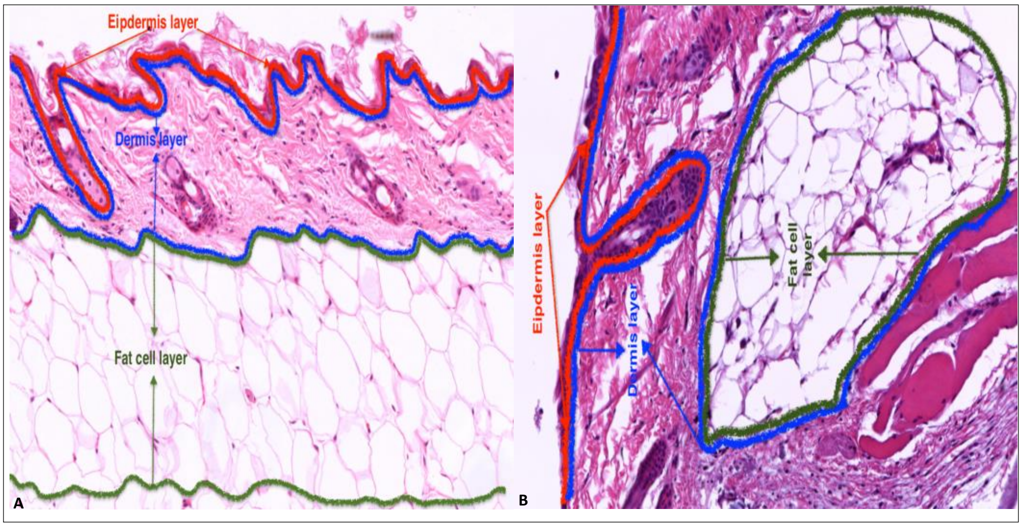


Figure 1.2 Different ROIs in mouse skin layer. (A, B) Epidermis ROI (red line), dermis ROI (blue line), fat cell ROI (green line)

5. Wild type samples. The aim of this research is to find the disorder of the microscopic skin images, to do that it need to have two types of samples, wild and mutant type. Our data set is 7000 H&E images divided to 6500 as wild type and 500 images as mutant type. Where those numbers of wild samples is not enough to perform an accurate analyse to extract features of them.
6. Damage samples. In the our data set there are many damage samples of H&E images in the wild and mutant types. Where the accuracy of our results was affected by the damage samples, because they effect directly in the accuracy of the segmentation and quantification in different layers. For example, it was damage in layers segmentation and nuclei quantification in several layers as shown in Figure 1.3 (A, B, C).

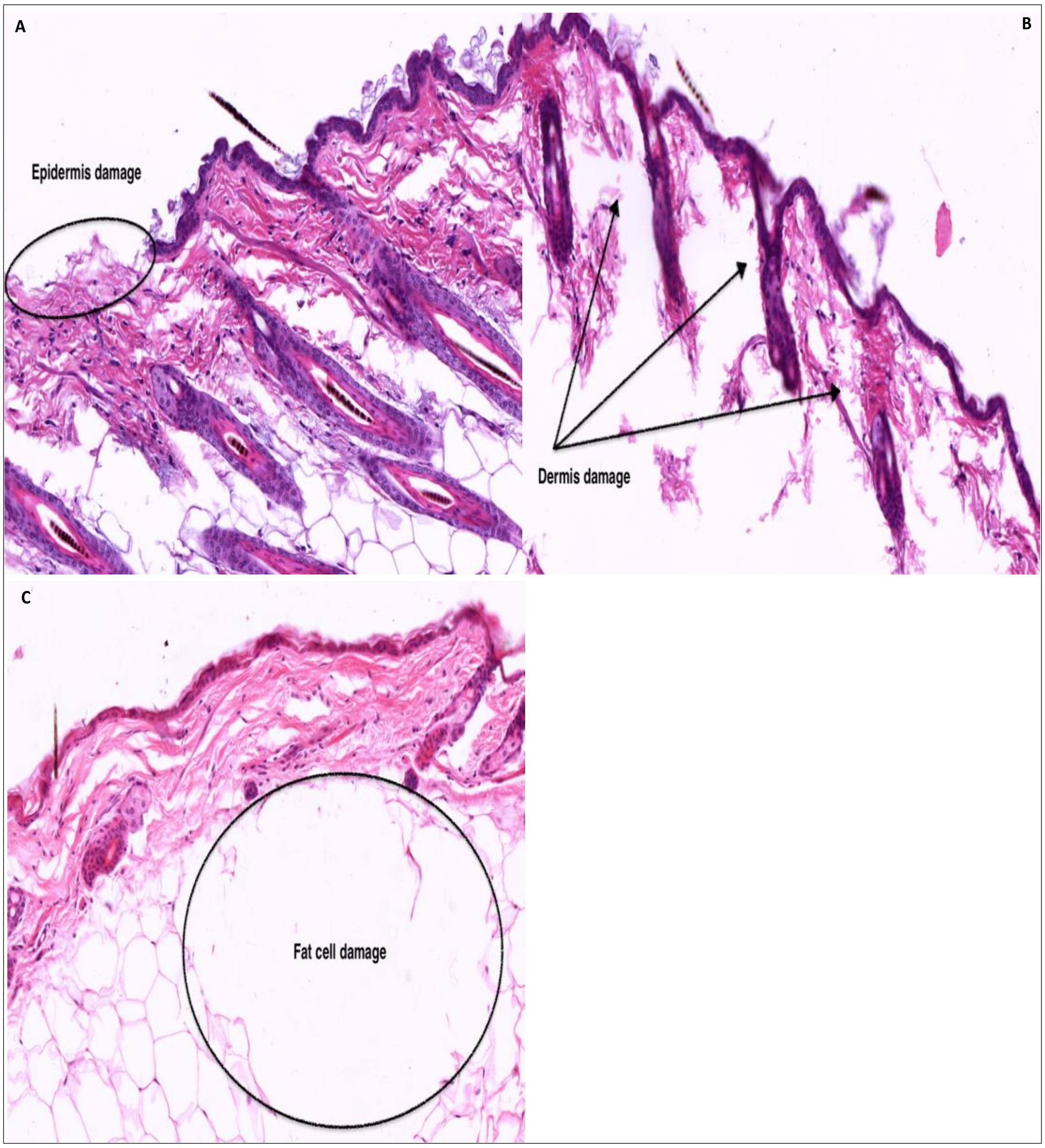


Figure 1.3 Damage samples of H&E images. A) Damage in epidermis layer. B) Damage in dermis layer. C) Damage in fat cell layer

1.4 Thesis Aim and objectives

Having identified the targets of our computational tasks, and the main challenges that we encountered during this project, we can state the overall aims and objectives of this research project as follows:

Overall aim. Develop a framework for a high throughput multiple image analysis tasks to facilitate the automatic annotating and quantification of skin abnormality related components/features suitable for integration into the Knockout genes research aspects of the MGP.

To fulfil this overall aim, we state the following items as the main objectives:

Objective 1. Develop and test the performance of automatic scheme(s) for microscopic skin image layers and sublayer segmentation.

Objective 2. Develop automatic scheme(s) for segmenting and counting nuclei in both the dermis and the epidermis layers, and sublayer segmentation. Test the performance of the developed schemes by comparing with the manual work of domain experts.

Objective 3. Develop automatic scheme(s) for segmenting hair follicles in the epidermis and determining their orientation with respect to a fixed alignment of the epidermis. Test the performance of the developed schemes by comparing with the manual work of domain experts.

Objective 4. Develop an efficient computational model of the human expert concept of epidermis curvature, and design a procedure to compute this parameter. Again, test the performance of this model by comparing with the manual work of domain experts.

Objective 5. Test the ability of using machine learning based on changes to the output parameters from each of the 3 previous automatic schemes, to determine genes known to cause skin abnormality/disease known to be associated with these parameters.

In all the first 4 objectives, the developed schemes are expected to overcome, or reduce the effect, of the challenges described in the last section.

1.5 The Contributions of the Thesis

1- Automatic layer and sub-layer segmentation of H&E mice skin images by fusing colour deconvolution method and fuzzy clustering.

2- The above fusion method relies on counting the histogram peaks of the images to improve the layers segmentation accuracy and overcome the effect of unintended staining errors.

3- The improved segmentation accuracy helps in proposing effective predictors of different sets of angles, features, distance and index-based measurements for quantifying hair follicles orientation, nuclei counting and epidermis curvature.

4- Developing simple, yet effective, methods to automatically align in a fixed image direction the epidermis layer in all H&E images.

5- Developing a hair follicles detection and quantification of their orientation.

6- Automatic detection and counting of nuclei in epidermis and dermis layers.

7- Develop a simple computational scheme for the automatic quantification of a parameter that models what experts interpret as the epidermis layer curvature.

8- And finally, identifying a list of candidate genes that might be responsible for abnormal changes in the numerical outputs of the last 3 schemes relating to nuclei counts in the epidermis and dermis layers, hair follicles orientation, and curvature of the mice skin layers.

Besides, we anticipate that eventually, the analytical techniques developed in this thesis, could have a wider implication once prepared in formats suitable for integration into clinical systems diagnosis and treatment assessments. The novel combinations of image processing/analysis techniques and pipelines will save time and are expected to produce more reliable results that will ultimately help or improve the speed and quality of dermatology and cosmetic treatments.

Contribution points 1 and 7 have led to the following publications:

1- **Saif Hussein**, Joanne Selway, Sabah A. Jassim, Hisham Al-Assam, "Automatic layer segmentation of H&E microscopic images of mice skin", in Mobile Multimedia/Image Processing, Security, and Applications 2016, (SPIE, Bellingham, WA 2016), 98690C.

2- **Hussein, Saif** and Jassim, Sabah A. and Al-Assam, Hisham (2017) Automatic Quantification of Epidermis Curvature in H&E Stained Microscopic Skin Image of Mice. In: Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science, 723. Springer International Publishing, Cham, pp. 935-945. ISBN 978-3-319-60964-5.

1.6 Thesis Outlines

Based on the proposed aims of this thesis several works were carried out in a sequential manner. The thesis outline is organised as follows:

- Chapter **2** provides the biological and computational background to our research.
- Chapter **3** briefly describes our research framework and methodology the research process. The chapter also explains the experimental protocol and details about datasets and ground truth that are used in this study. The chapter serves as a roadmap for the research works presented in the following key chapters.
- Chapter **4** describe the first contribution of this thesis towards layer segmentation method. Firstly, developed method based on colour deconvolution was proposed to automatically segment the three main layers of mice skin. Then an unsupervised segmentation method based on an improved the layer segmentation was proposed to segment the layers. Over all contribution in this chapter was proposing the fusion segmentation method, which merge two methods and chose the appropriate method based on the peaks of image.
- Chapter **5** presents second contribution, which is to segment the nuclei in epidermis and dermis layers automatically based on unsupervised technique.

This chapter consist of two methods for segmenting the nuclei, automatic method for nuclei segmentation in the epidermis layer, which include automatic nuclei separation based on features extractions and distance transform associated with watershed. Then another method has proposed in this chapter, which was nuclei segmentation in the dermis layer based on unsupervised technique, and improvement was introduced based on the shape feature extraction.

- Chapter **6** provides another contribution was done in this thesis, which was quantifying the orientation of the hair follicle in the epidermis layer. To address the main issue in this chapter, there are two automatic methods was proposed. First method was automatic alignment method of the epidermis layer. Second method was automatic segmentation of the hair follicle in the epidermis layer. Finally, based on two proposed methods, the main contribution in this chapter was proposed to measure the orientation of the hair follicles.
- Chapter **7** describes the last contribution was quantifying the curvature of the epidermis layer. Where the distance for each point between the reference line and the epidermis was used to measure the changes in the epidermis curvature.
- Chapter **8** summarises the main conclusions obtained in this research. In addition, the suggestions for further study are drawn in the end.

Chapter 2 : Biological and Computational Background

This chapter is devoted to briefly describing the biological and technical backgrounds for this research project. It is not about providing a comprehensive review of the state-of-art literature on image processing techniques for problems relevant to this research. This chapter could be skipped by readers of the relevant biological or technical backgrounds.

The technical background part of this chapter is therefore mainly concerned with image processing techniques for segmenting images.

2.1 Biological Backgrounds

This part of the chapter aims to provide a basic understanding about the biological relevance of microscopic image annotations to be achieved through this research. This part mainly describes the mouse genome projects and highlights their main findings and importance for human health. I also explain the justification for annotating a huge volume of H&E images of mice, and hence the need for specially developed automated solutions for the annotation and the desire of discovering any strong links between the automatic annotation results and classification of abnormality of skin features as results of gene knockouts.

A genome is an organisms complete set of DNA codes that includes all its genes, and provide the information required to construct and maintain the organism. Each cell in the human that has a nucleus that keeps a copy of the entire genome (Genetics Home Reference, 2003). It has been confirmed that some rare diseases happen because a single gene stops working (*Genes in Life | Genetics 101*, 2016). Due to such a close relation between genes and the diseases.

Scientists hope to use genetic information to help in diagnosing, curing and may be preventing such diseases. Two genome projects were initiated to define the gene sequences functionalities and their relations to diseases and abnormalities. The first one is MGP (Wellcome trust Sanger Institute, 2016) and the second is HGP (National Human Genome Researche Institute, 2015).

2.1.1 The Human Genome Project

The well-known international Human Genome Project (HGP) that was completed over the period 1990-2003, aimed to define the human genome series and recognize the genes it contains. It produced a full sequence for 3 billion DNA base pairs, and recorded all the human genes of the 25,000 (Genetics Home Reference, 2003). The most successful common scientific accomplishment was finding the genes sequence. The final project enables scientists to find accurately any problems in a single gene by applying a huge number of tests of available gene (Chong *et al.*, 2015).

Despite the success of HGP, the long-term ambitious goal of determining the functionality of all genes and its relation to human diseases are yet to be reached. This requires launching different projects that attempt to model human diseases in terms of genetic information, while addressing ethical and legal concerns.

2.1.2 The Mouse Genome Project

To achieve clear understanding about gene functionalities in relation to diseases, the mouse has been chosen as a mammalian model system for genetic research. This is because of significant similarities between the human and mouse in terms of genetic and physiologic makeup (von Scheidt *et al.*, 2017). The following facts justifies this project:

1- There are similarities between humans and mice in physiology, development, illnesses and body plan.
2- Mice are small in size, and the large litters ease the maintenance of a large number.
3- 95% of human genes are similar to those of the genome of mice (von Scheidt et al., 2017).
4- Availability of techniques to change gene expression accurately and efficiently (Sundberg, 1994; Nakamura et al., 2013; Severin et al., 2017).

In 2011, the International Mouse Phenotyping Consortium (IMPC) was launched as a 10-years project which aims to discover the functionality for every mouse gene by generating and systematically phenotyping 20000 knockout mouse strains (IMPC, n.d.). The first stage of the project aims to produce and phenotype 5000 mouse lines from 2011 to 2016 while the rest 15000 mouse lines should be completed by the 2021 (Adissu *et al.*, 2014). Before the IMPC project, the WTSI started the MGP in 2006 as a one of the first program for large-scale gene function analysis in mammalian system (White, Gerdin, Karp, Ryder, Buljan, Bussell, Salisbury, Clare, Ingham, Podrini and others, 2013). Later, the Sanger Institute become one of the founding members of the IMPC. Manipulation of the mouse genome involves a process called gene knock out (a biological process by which an existing copy of a gene is knocked out and replaced with an artificial piece of DNA). The inactivated gene will lead to change in a mouse's phenotype causing specific behavioural and other biochemical and physical characteristic changes (National Human Genome Researche Institute, 2015).

2.1.3 Wellcome Trust Mouse Genetics Project

Mice genetic background and phenotype data is generated by Wellcome Trust Sanger Institute (WTSI), for the benefit of the entire scientific community (Wellcome trust Sanger Institute, 2016). Our collaboration with WTSI focuses on the phenotyping of a subset of genetically knockout mice. The program at WTSI to functionally annotate the mouse genome is illustrated Figure 2.1.

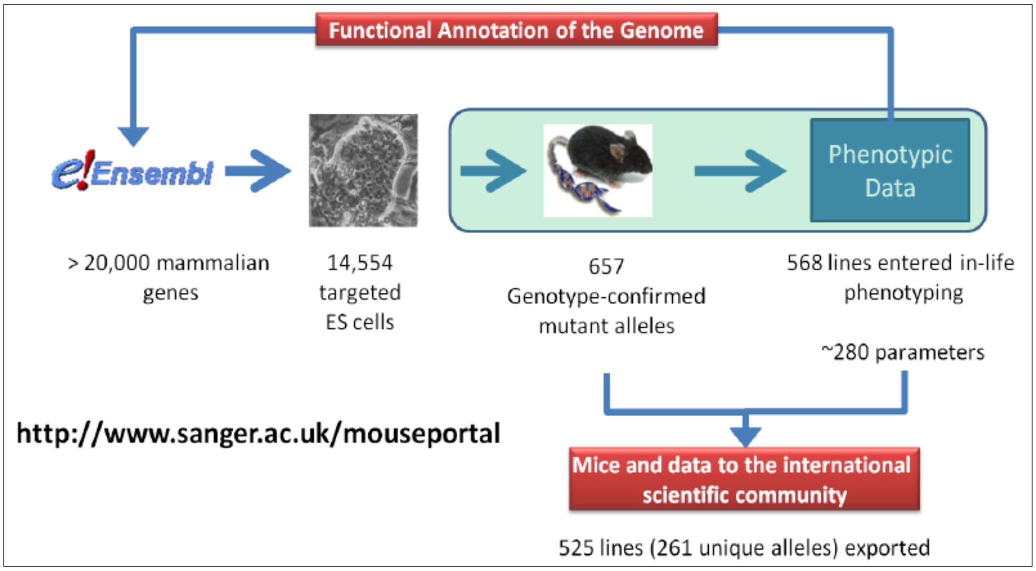


Figure 2.1 Overview of the functional annotation process by the Mouse Genome Project (MGP) at the start of this project. Many more lines are now available to study. Image courtesy of the WTSI (Wellcome trust Sanger Institute, 2016)

The primary phenotype data from WTSI aims to identify novel genes which related with diseases, such as cancer and developmental genetics. There are other research projects that focus on the functions of genes in the mouse genetics area by WTSI, relevant to mouse behaviours, cancer and developmental genetics (Liakath-Ali *et al.*, 2014). Access Data is available via the mouse resources portal (http://www.sanger.ac.uk/mouseportal/).

2.1.4 Secondary Phenotyping from WTSI Resources

Secondary phenotyping is about the detection of interesting features in gene-knockout tissue sections to detect early phenotypes. One example of secondary phenotyping is the assessment of genetic determinants of bone mass and strength, which can be involved in the pathogenesis of osteoporosis (Brommage, Powell and Vogel, 2019). Using high throughput investigation of mouse skeletons, biomechanical testing and statistical analysis of 100 mice generated by the MGP pipeline, nine new genetic determinants of bone mass and strength were found (Liakath-Ali *et al.*, 2014; Brommage, Powell and Vogel, 2019). In this thesis, we are interested in the assessment of genetics determinants of skin tissue in relation to abnormalities and skin diseases.

2.2 Skin Structure and Skin Histopathology

Mammalian skin tissue consists of three layers: Firstly, the external layer is the highly organised epidermis, secondly the dermis (connective tissue), and finally adipose layer (the subcutis) as illustrated in Figure 2.2.

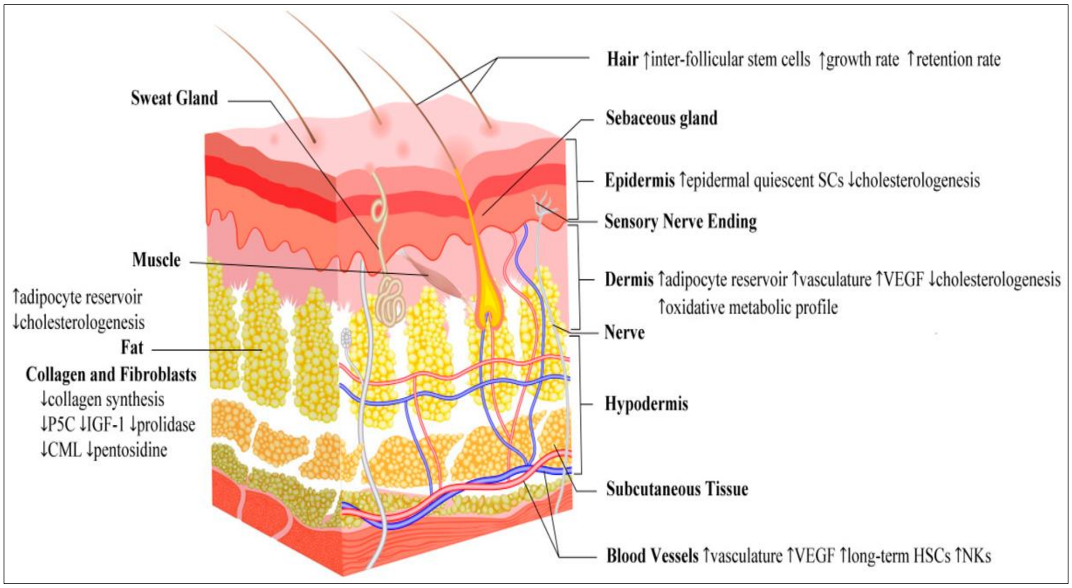


Figure 2.2 Structure of skin layer (Bragazzi et al., 2019)

2.2.1 Skin Function

The most obvious function of skin is to maintain a barrier that prevents the loss of fluids and protects against penetration by micro-organisms and UV radiation (Goldman, 2008).

However the skin provides several other functions: it regulates body temperature by blood flow and limits the inward and outward passage of water. The secretion of sweat and lipids causes the elimination of a number of harmful substances resulting from metabolic activity. The skin also provides flexibility and strength to protect the body from friction and impact wounds. Also, the skin can act as a sensory organ as it has a large number of nerve fibres and nerve endings. The skin also produces vitamin D when it is exposed to the sun, which is an important substance in bone health (Gherardi, 2008; Gawkrodger and Ardern-Jones, 2016). Where the skin consists of three layers (epidermis, dermis and fat cell) and each layer contains objects such as the hair follicle in the epidermis and the nuclei in different layers.

2.2.2 The Epidermis

The epidermis is the external layer, which act as chemical and physical barrier between the exterior perimeter and the interior of the body (Plotczyk and Higgins, 2019). And is consisting of Merkel cells, melanocytes and Langerhans cells, and it is include of layers of keratinocytes. Four layer in the epidermis: corneum, spinosum, basal and granulosum as shown in Figure 2.3.

Protect the body from several stressors environmental were the main purpose of epidermis layer, such as water loss, damage injuries and infections. The sweat glands in the dermis layer and the hair follicles in the epidermis layer were a good example for continuous interactions between dermis and epidermis layer.

Epidermal thickness depends on body site, e.g. on the eyelids it is about 0.05 mm, but on the soles of the feet and the palms of the hand (plantar skin), it is 0.8±1.5 mm (Gawkrodger and Ardern-Jones, 2016). In thick epidermis, an additional layer called lucidum can be seen, which represents a transition between the granulosum and the corneum (Bigby, 2011).

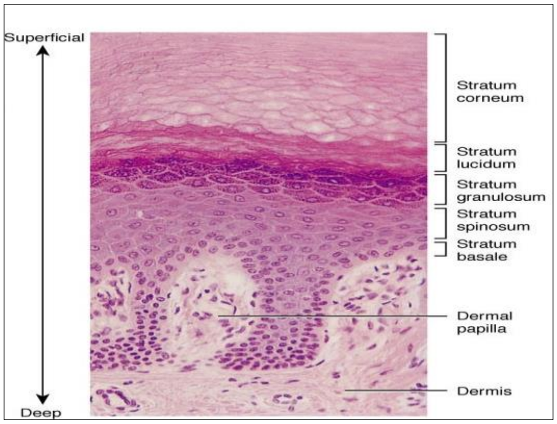


Figure 2.3 Upper layers of the skin. The four epidermal layers with dermis (Goldman, 2008)

2.2.2.1 Existing methods

Traditional skin studies utilize manual methods for the quantification of skin feature parameters. Recently, there has been a move towards automating the quantification of different skin associated features in relation to skin diseases, in order to increase efficiency by reducing processing time, improve accuracy by reducing human error, and by reducing laboratory costs. We aim to demonstrate that quantifying changes in epidermis curvature is of some scientific interest and clinical relevance. Unfortunately, our literature review did not unearthed directly related algorithms. We shall review existing methods on quantifying other types of skin features/objects and their linkage to skin disorder in certain diseases.

Epidermis layer was segmented automatically in (Lu and Mandal, 2015) as part of a diagnosis and analysis system to identify skin cancer particularly melanoma. Pathologists stipulate that important cues about melanoma are concentrated in specific skin areas (such as the entire epidermis, keratinocytes and sample of melanocytes), and hence priority is given to segmenting these objects. Discriminating features, extracted from these objects, with the support vector machine are used for classification.

Measuring melanoma tumor parameters in microscopic images of the epidermis was addressed in (Xu *et al.*, 2017) by proposing a four steps method: (1) Mahalanobis distances associated with colour features is used to identify melanoma regions, (2) multi-thresholds was employed to segment the epidermis, (3) Bayesian classification was applied to identify the skin granular sub-layer of the epidermis, and (4) multi-level of approach is used to measure melanoma area by utilizing the Hausdorff distance, by calculation of mismatch distance between epidermis layer and melanoma depth of invasion.

For treatment of melanoma, the depth of the tumor is a very important factor in identifying and targeting the tumor accurately. Depth measurement of tumor in epidermis was investigated by (Noroozi and Zakerolhosseini, 2015). This method has many stages, which start with segmenting the cornified layer by applying the entropy feature of the ridges in the images. Then active contour associated with colour features is used to segment the cells nuclei in epidermis. Finally Fourier feature extraction with index filter is applied to recognise the Melanocytes from specimen lymphocytes.

Measuring epidermis depth was also investigated in (Xu and Mandal, 2015a) using a set of steps. The method begins with segmenting the epidermis by employing shape feature and Otsu's thresholding. Then thickness of epidermis layer is calculated by using Euclidean distance of intersection points of perpendicular line which is connect the boundary of epidermis (top and bottom sides) Next, the method able to recognize the inaccurate segmentation by matching the extracted thickness with threshold, the value of the thickness must be smaller than or equal the obtained threshold to be accurate, otherwise it will be inaccurate segmentation. Ultimately segmentation refinement was done utilizing k-mean techniques to improve the accuracy.

Note that quantifying the follicle orientation, in the last chapter, relied on using an appropriate simple method to determine the two bounding curves of the outer component of the epidermis.

In relation to our objectives, the relevant solutions above for epidermis related quantification are those that quantifying epidermis thickness measurement, but none attempt to quantify epidermis curvature measurement. This may be attributed to the fact that, unlike our work, these research were not aiming to investigate the genetics causes of skin abnormalities and skin disorders diagnostics. Moreover, some of these relevant work do not support segmentation of the epidermis sub-layers, which is an important stage to have accurate curvature representation.

2.2.3 The Dermis

The dermis is consider as the main part of the skin, where a thickness of about 1- 3 mm. Dermis has many functions such as mechanical resistance and flexibility to the skin, interacting with outer layer (epidermis) to protect and maintain the dermo-epidermal junction and keep save pilosebaceous units, sweat glands and nerves (Jensen and Proksch 2009),(Haake *et al.*, 2001).

2.2.4 The Subcutis (or Hypodermis)

The inner layer of the skin is the subcutis.it include a clear fat cells and connective tissue, where it lies down the dermis. And it is also known as the adipose tissue layer. Human body include several of fat depots, where the largest fatty deposits are lie downward the skin in the human body, and surround fat in the organs in the body (visceral fat) (Burns, 2010). The thickness of the subcutis varies in different parts of the body: this layer can be up to 3 cm thick on the human abdomen. The striated muscles separate the Subcutaneous fat from the rest of the body (Burns, 2010).

There are two type of adipocyte in the human body, white and brown adipocyte, which have different properties, for example white adipocytes have less cytoplasm due to the large fat store within the cell. Brown adipocytes have a large amount of cytoplasm with a centrally located nucleus and smaller pockets of fat stores within each cell. The location and function of white and brown fat in the human body is different. White adipocytes function as energy stores and brown adipocytes function as energy consumers. Normal skin contains only white adipocytes, however, the size and number of adipocytes can change in different pathological states, so a decrease or increase in fat can indicate an abnormality e.g. obesity or diabetes (Britannica, 2018).

2.2.5 Dermal-epidermal Interface (Dermoepidermal Junction)

The boundary between the epidermis and dermis include of a specialized aggregation of attachment molecules and connective tissue, collectively known as the basement membrane (Gawkrodger and Ardern-Jones, 2016). This structure is complex and is of considerable interest as genetic defects in its composition leads to a variety of diseases, and it also serves as a target of autoimmune attack (Goldman, 2008). The dermoepidermal junction flattens during ageing (Farage *et al.*, 2010), which accounts in part for some of the visual signs of ageing.

2.2.5.1 Existing Solutions on Skin Layers Segmentation

There are a number of simple segmentation techniques, such as a threshold on image intensity or edge detection, that have been used for segmenting skin layers. The simplest method to segment an image into different regions is the use of a global threshold, especially if the regions are different enough in terms of intensity values to allow the separation of the object and the background (N and S, 2016).

Colour skin image of human face was segmented by using HSV (hue, saturation, value) in (Zainuddin, Naji and Al-Jaafar, 2010), and succeeded in greatly reducing false negative cases by employing colour space technique to cluster colour. Four colours were clustered in the face skin images: standard-skin style, light-skin style, redness-skin style and shadow-skin style. The 3D colour space was then transformed to 2-D to perform the information. Pixels and regions based segmentation were essential stage to improve the accuracy.

In another study, the epidermis layer was segmented using microscopic H&E skin images based on threshold and shape in(Xu and Mandal, 2015b). Then epidermis thickness is calculated by measuring the distance between main axis of image and groups line perpendicular of epidermis segmented mask. After that the k-means method is used to fine segmentation to have accurate segmentation. Many methods were used in the previous examples that are nor suitable for our work. First, none accurately segmented the main mice skin layers, and second none used a connected mask to segment the mice skin layers, which is essential for our work.

Epidermal segmentation has been evaluated manually by measuring the thickness in (Sandby-Moller, Poulsen and Wulf, 2003), and the relationship between the thickness of the stratum corneum and the cellular epidermis was investigated to find any correlation between age, gender, body site, pigmentation, blood content, smoking history and skin type. Biopsies were taken from different areas such as the dorsal forearm, shoulder and upper quadrant of the buttock. The results showed a significant influence of these factors on epidermal thickness segmentation. Although this study provided objective measurements, the number of subjects was not sufficient, and a limited number of parameters were measured. This study, nevertheless, reinforces the need for automated unsupervised techniques to assess skin in high throughput histopathological analyses.

Another interesting work for segmentation is based on using active contour method to separate the layer of mice skin in (Osman *et al.*, 2013). Prior to that, some pre-processing operation has been done, such as contrast enhancement to highlight the objects, that are not prominent in the image, and then the median filter was applying to reduce the noise. Subsequently, the image is added to itself, by adding intensity values for each pixel in the image with the corresponding pixel intensity in the same image. This helps produce an output image more suitable to detect the epidermis in the image, by turning the epidermis region darker than other regions. Next the image is converted from RGB colour to HSV colour space, while the saturation channel was selected for the segmentation process because the region of interest (epidermis and dermis) contained saturated colours in the HSV colour space. Then Otsu threshold method has been applied on the saturation image to segment into two components: white pixels being the region of the epidermis and dermis, and black pixels being the background. The largest object is selected from the binary image that represents the initial curve. Finally, the morphology operator has been applied using “open” operation, to make the initial curve as an input to active contour algorithm, which covers the whole area of the dermis and epidermis in the original image.

Being the nearest to what we need, we implemented the above algorithm and found that the active contour method was appropriate method for their own objective of segmenting the adipocyte cells in the fat cell layer. These cells are immediately adjacent to the epidermis and dermis, so by isolating these two layers and using subtraction of the segmented epidermis and dermis from the saturation image, quantifying the adipocyte cells became straightforward.

However, the algorithm does not perform well enough to segment other layers or to segment sublayers of the main layers. In particular, we found that this algorithm fails to segment the two sub-layers of the epidermis, the basal and the cornified layer. The following example illustrate in Figure (2.4) this shortcoming of the above algorithm.

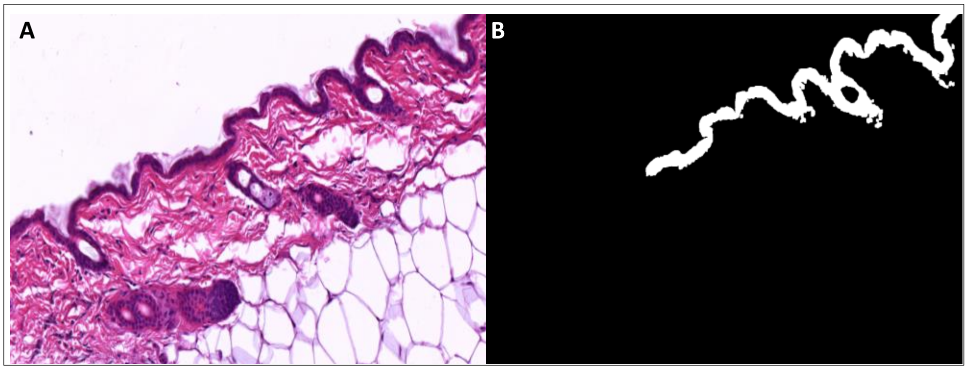


Figure 2.4 Output of Osman's method segmentation. A) Input colour image. B) Output segmentation

Since this sub-layer segmentation is essential for our main investigations on abnormalities in the orientation of hair follicles and the number of nuclei in different layers, then we need to develop a more sophisticated multi-layers and sublayer segmentation schemes.

2.2.6 Hair Follicle

The hair follicle is a unique structure (Figure 2.5) that is constantly undergoing growth (anagen) period, regression (catagen), resting (telogen) and shedding (exogen) throughout the lifetime of an individual. Bulge stem cells, located at the insertion of the arrector pili muscle (APM), give rise to progenitor cells that proliferate during anagen producing a downward expansion of outer root sheath (ORS) encasing a specialised area of mesenchyme called the dermal papilla, together forming the hair bulb. Within the hair bulb are rapidly dividing matrix cells that produce the hair shaft and inner root sheath (IRS). The hair shaft and IRS are formed in distinct and concentric layers, analogous to onion skin, in which keratin proteins are prominently, expressed resulting in hair keratinisation and the formation of the rigid hair structure (Paus and Cotsarelis, 1999; Stenn and Paus, 2001; Schneider, Schmidt-Ullrich and Paus, 2009).

Hair follicles development occurs during embryogenesis when signals from the dermis stimulate the epidermis to produce a thickening of columnar cells, called the placode. Signals from the placode then induce formation of a dermal condensate below the placode. Down growth of the placode ensues encasing the dermal condensate, eventually forming the dermal papilla. Epithelial cells ultimately differentiate into specific layers resulting in hair shaft generation. Appearance of Wnt/β catinin signalling is crucial for initial hair follicle development (Paus and Cotsarelis, 1999; Stenn and Paus, 2001). Hair follicles may respond to inflammatory damage in a limited number of ways depending on the type, location, intensity and duration of the inflammatory attack. Premature catagen induction, dystrophic anagen, programmed organ deletion. Sometimes the response permits ongoing (although imperfect) hair shaft growth or a temporary shift in the hair cycle to allow time for follicle recovery. However, other times permanent hair follicle loss will ensue (Hadshiew *et al.*, 2004).

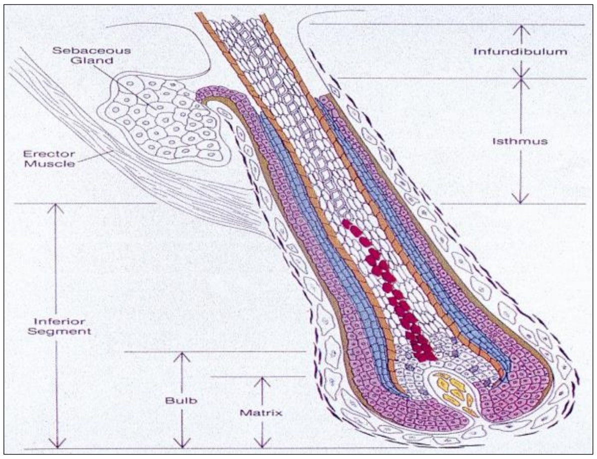


Figure 2.5 Hair follicle anatomies (Harries et al., 2009)

The skin image displays significant information about colour across images and the spatial arrangement of changes in intensity. The disorder in growth development of hair follicle in terms of colour and size in the skin surface, as well as the disruption of orientation in the epidermis layer, is of great interest to analyse skin abnormalities (DiTommaso *et al.*, 2014), (Scott, 2015).

Unknown functionality or lack-of-function of hair phenotypes in physical mouse mutants were the main reasons to study the phenotypes of hair. Previous studies have provided a helpful mechanistic view to show that causes of disorders of human hair growth are largely unknown. For instance, such reasons could include a set of genodermatoses (hypotrichosis simplex, monilithrex) or due to immunological conditions (e.g. alopecia areata), as well as hormone-dependent distortions, hair shaft disorders, hirsutism and hypertrichosis. The skin of mutant mice contributes credible evidence for identifying mechanisms of disease, which is comparable to the human structure (Sundberg and King, 1996; Nakamura *et al.*, 2013; Scott, 2015).

2.2.6.1 Existing Methods of determining Hair Follicle Orientation

In previous studies such as (Guo, Hawkins and Nathans, 2004; Wang, Badea and Nathans, 2006; Devenport and Fuchs, 2008; Ravni *et al.*, 2009; Wang, Chang and Nathans, 2010), large number of hair follicles has been quantified, in terms of the 2D structure of skin tissue. Furthermore, the WT and Fz6-1 backgrounds were assessed, according to the relative orientations of follicle-associated structures and the orientations of hair follicles, while in the mutant type of mice in Frizzled (Fz6) or Celsr1 and Vangl2 lead to disorientation of hair follicle.

The above mentioned studies employ different methods, guided by genetic information, to quantify the hair follicle orientation in an image. The adopted methods depend on the planar cell polarity to determine the angles of hair follicle relative to the skin surface, and then assess changes in these angles to match with the phenotype depend on knock out genes (Chang and Nathans, 2013; Ahtiainen *et al.*, 2014; Chang *et al.*, 2015; Chang, Philip M. Smallwood, *et al.*, 2016; Cetera *et al.*, 2018).

Inactive Frizzled6 (Fz6) gene in mice disruptes the hair orientation in (Wang, Badea and Nathans, 2006), which results in groups of hundreds hairs appear as waves in some area over the skin. By analyzing hair follicles growth orientation, it is found the orientation of WT grow in a particular order, while in the mutant mice the grown hair follicle orientation appear in random pattern. The process of determined automated model to quantify the orientation of follicle is done essentially based on the average of neighbors orientation of hair follicles constrained at a certain time. In mammals, hair follicles are associated with many objects such as: Merkel cells, arrestor pili muscles (APMs), cutaneous sensory nerves and sebaceous glands, each of which connect with the hair follicle from different directions. Relying on the association of each of these objects provide a different method to quantify the orientation of the hair follicle. In this work (Wang, Badea and Nathans, 2006), the follicle angles have been measured according to four associated objects as mentioned earlier. The most interesting issue was measuring the angles according the subcutaneous glands. Although this study provides objective measurements for hair follicles in the skin, the method used was not perfect, because it was still a manual system which is time consuming and not cost-effective. This study, therefore, reinforces a need for the use of automated, unsupervised techniques to assess skin in high throughput histopathological analyses.

2.2.7 Nuclei in the Skin

The nucleus is the site of the storage, which has genetic information that responding to the signals cell. Nuclei are changes in the condition such as skin cancer.
(Woodcock and Ghosh, 2010).

2.2.7.1 Relevant Work

Nuclei are deemed to influence the cellular rate of metabolism. In (Yang, Li and Zhou, 2006), it was shown that changes in the number of nuclei in skin tissue are one of the important skin features that reflect changes in metabolism i.e. an increase in number correlates with metabolic diseases such as cancer.

The advantage of the computerised method, over manual counting of nuclei in skin layers, is that it can conduct the task for large number of segmented epidermal and dermal tissues in a short time, and free from observer bias. Moreover, the images can be stored and used again as future reference. While in the manual method when the stain has been used on an image, it highlights some features in the image tissue, the rest of the details depend on the use of a stain. These observations explain the strong research interest in general automatic image segmentation. Next paragraph review relevant work on segmentation of nuclei like objects and discuss their effectiveness.

In general, image segmentation method rely on using a variety of image features/information to partition images into visibly different regions containing interesting details. Existing nuclei segmentation techniques combine the use of colour thresholds and/or edge detection. The use of colour thresholds alone is not expected to be effective enough to segment the nuclei in the epidermis layers in H&E stained skin sections because of homogeneity of colour intensities in the overall image, which might affect the thresholding process.

The use of edge detection for segmentation, implicitly assumes that image objects of interest have distinct boundary components separating them from background or other objects. Such an approach was adopted by (Sadeghian *et al.*, 2009) to segment useful information in an image, such as nuclei in the epidermis layer, by applying edge detection filters to separate image objects from the background whereby the different disjoint image regions differ in terms of colour intensity.

The assumption that edges bound regions of image with closely related colour intensity is far from being true always. On the other hand, discovering regions of nearly uniform colour intensity maybe amenable to clustering image intensities. Hence, colour segmentation such as the work in (Rich and Whittaker, 2005; Cisneros *et al.*, 2011) is widely used to separate image content into different regions through clustering by the k-mean method. The K-mean clustering in colour segmentation overcomes problems associated with the widely used pixel intensity threshold alone. Success of colour based clustering algorithms, including the K-mean, is reliant on the image having sufficient variation of colour (Muthukannan and M, 2010). This is not guaranteed in our application, but we should nevertheless exploit any level of colour variation.

Figure 2.6 illustrates segmentation output results from existing methods, such as Canny edge detection (Figure 2.6C), colour segmentation using k-means clustering of L*a*b* colour space of the image (Figure 2.6B) (Muthukannan and M, 2010). The results of these techniques do not always individually yield effective segmentation of the nuclei in different skin layers, but each of these techniques may help with highlighting certain image features in other images stained with other staining methods.

A supervised machine learning cell segmentation algorithm was introduced in (Mao, Zhao and Tan, 2006) that uses a combination of two analysis algorithms. Firstly, an algorithm for marker detection was developed for the watershed to isolate touching cells or connected objects, such as overlapping nuclei. The classification of the output patterns relies on shape information and photometric data of the image, and was employed to produce a robust marker detection method. Secondly, the RGB image is transformed into the YIQ colour space (where Y is greyscale pixel value, I is the hue channel and Q is the saturation channel) for classifying pixels according to membership of the cells. The Otsu's threshold method is applied to the greyscale image (i.e. the Y channel) to binarise the image based on the histogram distribution of the pixel intensity. In this paper the touching cells were isolated successfully by using watershed method. The isolation of overlapping nuclei was one of our aim in Chapter 5, so we had used the watershed method to achieve this aim.

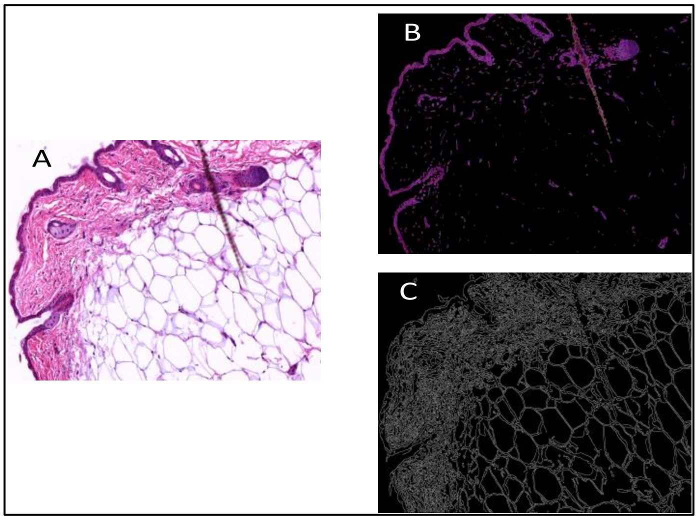


Figure 2.6 Nuclei segmentation with existing methods. A) Original H&E stained image. B) k-means clustering output image (k=3). C) Canny edge detection

2.3 Histology Staining Methods in Skin Analysis

Examining skin tissue slides under the microscope in terms of detection of its content objects require histological staining (Wei *et al.*, 2019). The use of special stains highlights features and different structures within a specimen. In terms of analysis, staining can also facilitate image segmentation. Several staining methods are used in histopathology for highlighting different features in skin tissue sections (Gherardi, 2008), and the main ones used in this study is explained below:

Haematoxylin & eosin (H&E) is a general stain. It can allow assessment of overall structure. Acidic substances (i.e. nuclei) are stained blue and alkaline substances (the majority of other tissues) stain shades of pink. Figure 2.7 illustrates a skin tissue section stained with H&E (provided by The Wellcome Trust Sanger Institute [WTSI] Mouse Genome Project [MGP], unless otherwise stated all other tissues were generated in the Buckingham Institute for Translational Medicine). This staining is widely used in morphological analysis, where each skin layer can be discriminated, which helps to measure the depth of layers.

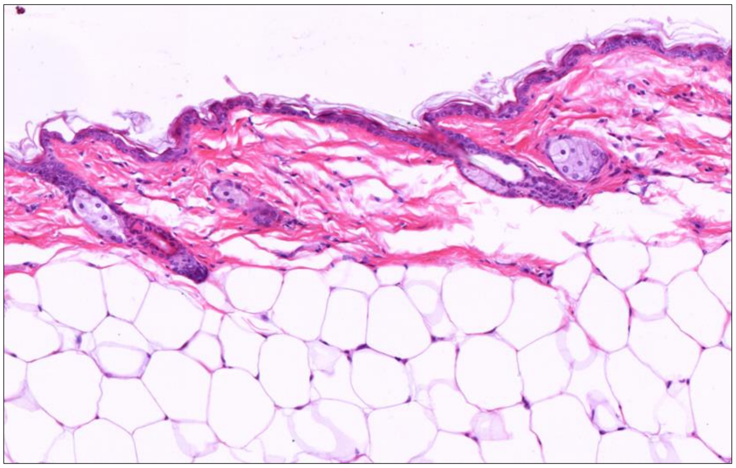


Figure 2.7 Skin image in H&E stain (Goldman, 2008)

2.3.1 H&E Staining of Skin

Using the H&E stained tissue, highlights its crucial features in several pathological condition. The need to show detail of huge information about the structure and organisation tissue was essential advantage to use the staining method (Bancroft and Gamble, 2008). Many projects have used the H&E stained histological sections, for instance the morphometric of skin layer, such as depth, shape, area and length (Bapure, 2012). Also a useful information can be reveal by using stain, such as hair follicle orientation in the skin, number and size (Bancroft and Gamble, 2008). In addition, the staining can help to use in assessing the extracellular matrix (ECM) and general structure of the dermis whence, structure and organisation for collagen (de Vries *et al.*, 2000). Customarily, manual methods are used to qualitatively assess in histological changes, this process is time consuming and the result follow to observer's view bias. An Automated techniques in this chapter have been develop to split a skin layer and sub-layer indifferent methods with identify and count the nuclei in the those layers, to determined early phenotypes of interesting in a high- throughput genetic screen.

2.4 Biological Image Processing and Analysis

Digital images, including the microscopic ones, encapsulate a huge amount of information about their content, and the interpretation of the information in these images need effective image analysis solutions. Image processing techniques are operators that transform the original image (input image) to produce an output image that is more suitable for analysis (Gonzalez, Woods and Eddins, 2004).

Computer vision incorporates different important functions for the quantification of image content information within using automatic and semi-automatic algorithms to enable various image analysis tasks with high levels of accuracy, speed and throughput.

Biomedical microscopic images normally require own pre-processing steps so that useful features can be extracted reliably. Depending on the type of problems, segmentation or detection of certain part of the image is often essential. Biological images come in many different forms, and even within the field of microscopy images can be acquired from many different types of microscopes and other devices such as whole-slide scanners. Extracting the complex information contained in these images is a demanding task, and manual methods tend to be inaccurate and not necessarily reproducible. Therefore, the biologist or physician increasingly needs automated image processing and analysis techniques to get an accurate result to reach an informed decision to help with diagnosis and appropriate management (Meijering and van Cappellen, 2006; Toennies, 2017).

MATLAB is a powerful scripting environment that readily allows the deployment of a broad range of image analysis operators, and is very powerful in prototyping novel analysis pipelines. Image analysis methods devised for extracting cutaneous features and analysis, are written in MATLAB 2014b.

2.4.1 Image Acquisition (Image Sources)

It possible to acquired Biological images from different sources, and most are turned into digital form, which may require enhancement prior to analysis. Before capturing biological images there are many processes to prepare the tissue or cells for photomicrography. These included fixation, sectioning and staining to preserve and highlight specific features of the section.

2.4.1.1 Microscopy

The human eye can distinguish wavelengths in the visible range of 400 to 750 nm (Rädler *et al.*, 2019). In addition, perceiving the contrast in an image or object depends on the variance in intensity for two different regions (A1, A2) in an image, given by

$$Contrast = \log10(A1, A2) \qquad (eq2.1)$$

The constraint in wavelength, contrast and resolution does the human view eye insensitive to find tiny objects, and unable to find some polarization states, where the resolution of human eyes is ~0.1 mm (Rittscher, Machiraju and Wong, 2008).

The microscope is contributed, to beat restraints of the human vision. It was several kinds of dissimilar microscopic techniques that one can employ to view small items in any tissue, also analysis any kind of structures.

Microscopic are based on sample types, which fall into three categories: single molecules, tissues and nuclei .It was several imaging mechanism that are utilized to develop dissimilar biological specimens; in this thesis the following methods are described:

- Fluorescence illumination: fitting for fluorescent specimen such as nuclei in tissue culture or fluorochrome-stained sections.
- Brightfield illumination: utilized to study stained tissue, naturally coloured specimens, fibres and hair. Polarized illumination: applied to study birefringent samples with a systematic structure such as thin mineral sections, hairs and fibres, bones and feathers.

2.4.1.2 Slide Scanners (Digital Pathology Platforms)

Visualising skin tissues generate huge images with very high-resolutions that are often obtained by using slide scanners, by performing line-scanning followed by image stitching. Scanners device as the Aperio ScanScope provide images of up to 40x optical magnification and allow “virtual microscopy” (Masson-Lecomte *et al.*, 2019). This technique has advantages in that it incorporates a huge amount of information in one file, and can show the entire tissue in one image that can be analyzed in one pass. However, the size of these images might slow down batch processing, and they cannot be easily analyzed by traditional image analysis techniques and require expensive commercial software, which requires expert optimisation and is generally not fast enough to give timely results.

Figure 2.8 shows the Aperio whole-slide scanner and a typical display of the software.

Figure 2.8 Aperio ScanScope S canner (Molin, Thorstenson and Lundström, 2014)

2.4.2 Image Processing

These stages start with segmenting the image into interesting partitions and extracting the features that require quantification and analysis. Object detection is a common biological image processing task that forms the second stage of automated bio-image analysis. Image pre-processing comprises a range of image processing techniques as described in the following sub-sections.

2.4.2.1 Related Work

The general approach of the proposed thesis is to apply the proposed methods on the H&E images from the WTSI MGP to enable investigation of the effect of different genes and identify the interesting genes.

Biological images contain a variety of objects, which may need to be extracted and quantified. Many of the existing tools do not cater for the array of objects found in the research or pathology context and there are difficulties in doing this in a time-efficient manner (Gauthier *et al.*, 2019). For example, there are few whole-slide application tools, and few resources for cutaneous research. Automated image analysis exists within the computational field, but is limited in biological applications (R. Dougherty and Shmulevich, 2012). Automation has two major advantages: firstly it prevents bias and secondly the process can applied to a large number of images(Kueh *et al.*, 2008). Biological images are often complex and contain noise that necessitates image processing to enhance the image and allow extraction of meaningful quantitative information. Attempts have been made to automate biological image analysis as following paragraphs.

Layer segmentation is an essential pre-processing stage prior to quantifying objects in different layers, such as segment nuclei in the epidermis and dermis layers. While the changes to these objects in numbers and shapes may be due to genetic mutations or skin disease such as melanoma diagnosis (Mokhtari *et al.*, 2014).

While developing a computer-aided system for the diagnosis of cervical intraepithelial neoplasia Wang et al. (Wang *et al.*, 2009), used the SVM machine learning to segment the relevant tissue layers.

Another system was developed by Mokhtari et al. (Mokhtari *et al.*, 2014) to segment the epidermal layer, for measuring melanoma depth of invasion in microscopic images. Morphological operations and global thresholding was used to segment the skin layers.

Lu et al. (Lu and Mandal, 2012) developed method to segment 40X magnification of H&E images different layers, based on shape analysis and global threshold. Where the k-mean cluster was used to improve the epidermis segmentation.

Segmentation of 10X magnified images were investigated by Haggerty et al. (Haggerty *et al.*, 2014), who was employing a contrast enhancement and thresholding based technique to segment the epidermis layer.

Automatic method for segmentation and counting a hair follicles in the skin has been developed based on edge detection and Otsu thresholding (Shih, 2015).

Parvin et al. (Parvin *et al.*, 2007) proposed a multi-pass voting technique for inferring the centres of cell nuclei based on statistical feature shape.

Qi et al. (Qi *et al.*, 2012) proposed a single-pass voting (SPV) technique for nuclei detection, which only performs one round voting followed by mean shift clustering to detect nuclei seeds.

The fuzzy cluster was used to detecting the nucleated red blood cells automatically in the (Das, Maiti and Chakraborty, 2018). In the ultrasound muscle images, the curvature has been quantified automatically based on wavelet method (Namburete, Rana and Wakeling, 2011).

Another interesting work for layer segmentation is based on using active contour method to separate the layer of mice skin in (Osman et al., 2013). All above work are support epidermis layer segmentation perfectly, but it doesn't support sub-segmentation of the epidermis layer into basal layer and cornified layer as we need in our research.

2.4.2.2 Image Segmentation

Segmentation is a process that is used to subdivide an image into different regions or objects. Segmentation depends on properties and structure in the image and the pixels related to the object region (Gonzalez, Woods and Eddins, 2004), biological image analysis and quantification needs objects to be detected and the features of objects identified. After segmentation the segmented region can be analysed then a meaning assigned to this region (Cisneros *et al.*, 2011), the inputs in the segmentation method are images and the outputs are attributes extracted from those images. Segmentation approaches are divided into two categories. The first category applies to colour images and the second category applies monochrome images (Hosea, Ranichandra and Rajagopal, 2011), segmentation of monochromatic images is based on properties of image intensity values such as discontinuity and similarity. The image segmentation techniques in different domains are explained next.

Colour Deconvolution Model-Based Segmentation

Colour deconvolution is one of the frequently used image processing function to separate overlapping spectra. It is useful in highlighting the region of interest and this method is suitable for H&E(haematoxylin & eosin) stained images (Ruifrok, Johnston and others, 2001), we had to use H&E images because all our datasets in H&E. And it is suitable to analysis microscopic images because it highlighting the region of interest such as nuclei in the epidermis layer. The image pixel values, representing the stain concentration, can be computed by measuring the optical density (OD) of the image to determine the highest pixel intensity in the specimen (Ruifrok, Johnston and others, 2001). In the haematoxylin & eosin (H&E), where the H channel (blue) highlights the epidermis layer, the majority of which consist of overlapping nuclei (Nava and Jaffe, 2005; Cuadros *et al.*, 2007; Kimm *et al.*, 2007), while the E channel (red) displays the texture of skin layers (Arvis *et al.*, 2011), as shown in Figure 2.9.

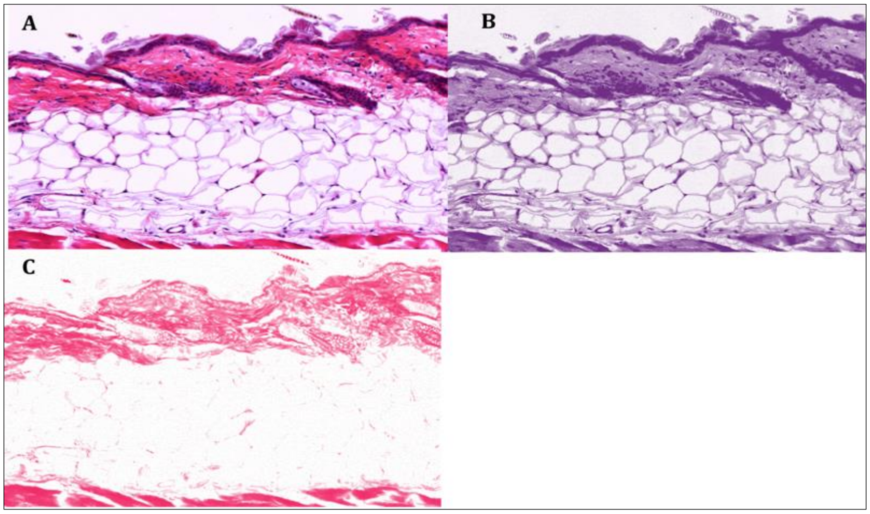


Figure 2.9 Colour deconvolution. A) Original H&E image. B) H channel. C) E channels

The intensities were detected from transmitted light (A) and (c) as absorption factor, by using the Lambert-Beer’s law (Peters *et al.*, 2019).

$$I_C=I_0, C \exp(-Ac) \tag{eq2.2}$$

Where $I_{0,c}$ represents light intensity, I_c is the light intensity after passing through a specimen, and c is the point to detection channel.

For individual channels of RGB the optical density (OD) can be defined as

$$ODC = -\log_{10}(I_C/I_0, c) = Ac \tag{ eq2.3}$$

It was evident that there was a clear linear relation between OD and every channel with intensity of absorbing material, so it can be capable of splitting several stains in a specimen.

Every stain is represented by optical density for the illumination in RGB, i.e. for each channel of the RGB image in pure stain term has an OD. For example, measure a stain with haematoxylin only, led to OD values of 0.08, 0.20 and 0.18 for B, G and R.

The vector will be proportional to the amount of stain, while OD will be calculated for each channel. In this case, there were three channels, while the relative values of the vector describe the actual OD for the detection channels. The matrix of colour can be described as following form.

$$\begin{matrix}[\boldsymbol{x11} & \boldsymbol{x12} & \boldsymbol{x13}] \\ [\boldsymbol{x21} & \boldsymbol{x22} & \boldsymbol{x23}] \\ [\boldsymbol{x31} & \boldsymbol{x32} & \boldsymbol{x33}]\end{matrix} \qquad \textit{(eq2.4)}$$

Each row relates to a different stain and each column represent the optical density, which was detected by the RGB channel for every stain. Measuring the absorption of RGB with a single stain led to the determination of the OD values for every channel. For example, the matrix of OD combination of hematoxylin (H), eosin (E) is:

$$\begin{array}{cccc} \mathbf{R} & \mathbf{G} & \mathbf{B} & \\ [\mathbf{0.18} & \mathbf{0.20} & \mathbf{0.08}] & \mathbf{H} \\ [\mathbf{0.01} & \mathbf{0.13} & \mathbf{0.01}] & \mathbf{E} \\ [\mathbf{0.10} & \mathbf{0.21} & \mathbf{0.29}] & \end{array} \qquad \textit{(eq2.5)}$$

To obtain separate information for every stain's contribution, the ortho-normal transformation of RGB needs to perform for partition of stains to obtain separate information for the contribution of respective stains. The transformation needs to be orthogonal and normalised to get a correct balance of absorption factor for each stain. For normalisation, the division operation taken place for every OD by its total length.

$$\hat{x}11 = x11/\sqrt{x_{11}^2 + x_{12}^2 + x_{13}^2} \qquad \textit{(eq2.6)}$$

$$\hat{x}31 = \hat{x}31/\sqrt{x_{31}^2 + x_{32}^2 + x_{33}^2} \qquad \textit{(eq2.7)}$$

$$\hat{x}21 = \hat{x}21/\sqrt{x_{21}^2 + x_{22}^2 + x_{23}^2} \qquad \textit{(eq2.8)}$$

Which leads to normalised OD array M:

$$\begin{bmatrix} \hat{x}11 & \hat{x}12 & \hat{x}13 \\ \hat{x}21 & \hat{x}22 & \hat{x}23 \\ \hat{x}31 & \hat{x}32 & \hat{x}33 \end{bmatrix} \qquad \textit{(eq2.9)}$$

The OD matrix M that was normalised as above merge of haematoxylin, eosin and DAB is:

$$\begin{bmatrix} \mathbf{0.65} & \mathbf{0.70} & \mathbf{0.29} \\ \mathbf{0.07} & \mathbf{0.99} & \mathbf{0.11} \\ \mathbf{0.27} & \mathbf{0.57} & \mathbf{0.78} \end{bmatrix} \qquad \textit{(eq2.10)}$$

If vector C is 3*1 for 3 stains at an individual pixel, then OD values at that pixel is y = CM.

It is clear from the above that C=M^{-1}[y]. In this sense, the multiplication operation between OD- image and inverse of OD matrix, that was defined as matrix D for the colour-deconvolution, which led to orthogonal representation.

$$C=D\,[y]. \qquad (eq2.11)$$

Matrix D is a colour deconvolution corresponding to matrix M for the haematoxylin, eosin and DAB matrix is:

$$\begin{bmatrix} 1.88 & -0.07 & -0.60 \\ -1.02 & 1.13 & -0.48 \\ -0.55 & -0.13 & 1.57 \end{bmatrix} \qquad (eq2.12)$$

In the resulting matrix, it can be observed that the diagonal elements of the matrix are larger than unity, while the negative elements are off the diagonal. To obtain the haematoxylin stain from colour from the above matrix (11), which indicates the OD (*eq2.4*) values for every stain in the image has to subtract a portion of the blue OD and the green OD from the red OD. To extract the eosin OD, a portion of the blue OD and the red OD must be subtracted from the green OD, while to get the DAB OD, a portion of the green OD and the red OD must be subtracted. In this status, the stains used will be pure green, blue and red stains, and the matrix above (11) will be the unity matrix.

Automatic classification of nuclei was contributed by using colour deconvolution in H&E stain images of three kinds of malignant lymphoma: mantle cell lymphoma, follicular lymphoma and chronic lymphocytic leukaemia (Orlov *et al.*, 2010).

The colour deconvolution method was used for the extracted haematoxylin stain, which represented nuclei in the colour image, which led to performing cell detection and segmentation by using a novel contour based minimum model (Wienert *et al.*, 2012; Irshad *et al.*, 2014).

A novel method has been introduced to normalise the stain in image of histopathology, such as tumour segmentation based on breast histopathology. Colour deconvolution was employed as a pre-processing stage of derived colour for nonlinear mapping image. The colour deconvolution method was a better method to obtain individual stain concentrations in a mixed stain image. Instead of using matrices of standard stain, which are probably inappropriate for the provided image, the proposed method utilises colour classifier, which is a new description of colour staining to determine the image matrix of a stain (Khan *et al.*, 2014).

Recent work of interest involved automated segmentation of microscopy stain images, which used the colour deconvolution method for obtaining the haematoxylin stain, in order to segment oropharyngeal cancer tissue micro-arrays (TMAs) and to identify stromal tissues and epithelium. Mathematical morphology was used for segmentation, while colour deconvolution was a main method to segment the haematoxylin stain from the H&E image, which focuses on the dark area of the nuclei. Unsupervised methods, the voting-based consensus function and Evidence Accumulation Clustering (EAC), were then applied to recognise stromal regions and layers (Fouad *et al.*, 2017).

Fuzzy C-Mean Model-Based Segmentation

The fuzzy c-mean method provided an accurate technique for crusting the image, by employing two values per pixel, which were the membership and distance. Based on these two values, it was determined in which cluster the pixel belongs. While the membership grade represents the degree to which the pixel belongs to each cluster, where the higher the value of membership indicates a higher probability of this pixel belonging to this cluster. In contrast, the low value of distance point to high probability of this pixel belonging to this cluster, as a result, each cluster consists of a pixel, which has a high membership grade with low distance (Babu *et al.*, 2010; Ng *et al.*, 2015).

The objective function was defined as follows:

$$J_{m}=\sum_{i=1}^{D}\sum_{j=1}^{N}\mu_{ij}{}^{m}\left\|x_{i}-C_{j}\right\|^{2} \qquad \textit{(eq2.13)}$$

Where

- m is any real value larger than 1
- Jm is the objective function.
- D is the number of data points.

N is the number of clusters.

- M is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$. Fuzzy overlap refers to how fuzzy the boundaries between clusters are, that is the number of data points that have significant membership in more than one cluster.
- X_i is the ith data point.
- C_j is the centre of the jth cluster.
- μ_{ij} is the degree of membership of xi in the jth cluster. For a given data point, xi, the sum of the membership values for all clusters is one.

The following steps show how FCM work in order:

1- Randomly initialize the cluster membership values, μ_{ij}.
2- Calculate the cluster centers:

$$\boldsymbol{C_{j}} = \frac{\sum_{i=1}^{D} \mu^{m}{}_{ij} X_i}{\sum_{i=1}^{D} \mu_{ij}} \qquad \textbf{\textit{(eq2.14)}}$$

3- Update μ_{ij}according to the following:

$$\boldsymbol{\mu_{ij}} = \frac{1}{\sum_{k=1}^{N} \left(\frac{\| X_i - C_j \|}{\| X_i - C_k \|} \right)^{\frac{2}{m-1}}} \qquad \textbf{\textit{(eq2.15)}}$$

4- Calculate the objective function, J_m.
5- Repeat steps 2–4 until J_m improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

The FCM method with orientation sensitive in H&E images is employed for automatic segmentation of the epidermal and dermal layers and other tissue, for instance melanocytic cells and nests which refer to the presence of cancer cells (Babu *et al.*, 2010). The proposed method was produced to assist the diagnosis function for dermato-pathologists. The automatic segmentation method technique of H&E stained images was created by applying the FCM for clustering the human skin tissue to find the region of interest which leads to determination of the abnormalities for human skin identified by medical practitioners when manually segmenting the ROI, (Mandal, Gupta and Kar, 2016).

Thresholding

The technique of thresholding is widely used in biological image segmentation because of the simplicity of implementation. This process results in a fast image analysis if the threshold value can be chosen automatically and adaptively to be applied to all images. The idea of a threshold is that the object and background in a given image f (x, y) have intensity levels grouped into two modes. These modes are separated by the threshold t in the intensity histogram of that image Figure 2.10 such that any point at which f (x, y) is greater than the threshold T is an object otherwise it is considers as a background (Dhara, Suryono and Widodo, 2017)(Esfahlani and Sayama, 2018).

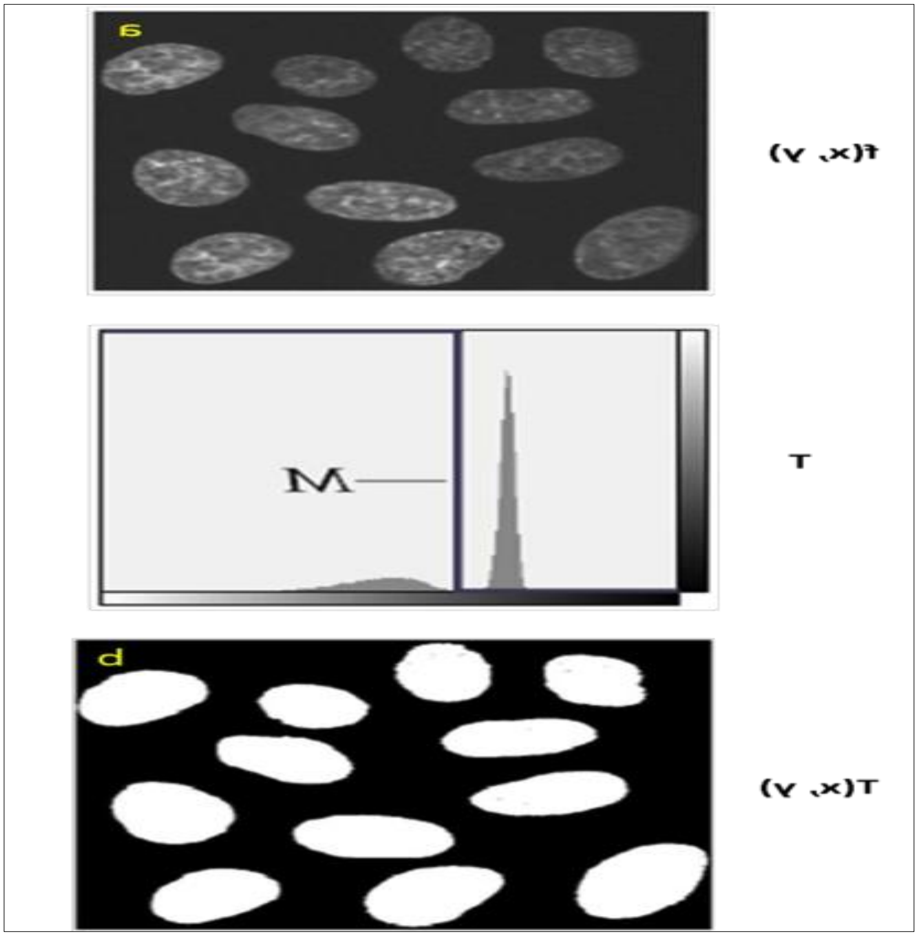


Figure 2.10 Thresholding. Thresholding achieved by applying the threshold function M to the input image f (x, y) and producing the thresholded output T (x, y) so that T (x, y)= M (f (x, y)). a) Grayscale image with the corresponding histogram underneath, b) Binary image

Euclidean Distance Transform

The distance transform is a method that usually only deals with binary images. Grey scale image is the result of the transformation with the same style of input image. The grey scale intensities of each point in the foreground area are changed to display the distance from each point to the nearest boundary. One of the most common distances is Euclidean distance transform, which finds the nearest straight line between two points, and it used to measure the points separation in the image, by determining the distance between the nearest nonzero pixel and each pixel in the binary image, as shown in *eq (2.16)* and in Figure 2.11.

Where

First point is $(A_1, B_1,)$, second point $(A_2, B_2,)$,

Euclidean distance = $\sqrt{(\boldsymbol{A_1}, -\boldsymbol{A_2})^2 + (\boldsymbol{B_1} - \boldsymbol{B_2})^2}$ (Maurer, Qi and Raghavan, 2003) ***(eq2.16)***

Figure 2.11 Two-dimensional Euclidean distance transforms

Watershed Transform

Watershed is one of most common methods for segmenting the medical image and other types of images to extract the particular object from the image. Topographic surface on the grey scale mode is made by the algorithm of the watershed, where the heights is technique to represent the intensity of pixels. The catchment basin of the image is the local minima region, where the water flooding starts and increases until the surrounding area floods. The boundary between the two basins is a watershed as shown in Figure 2.12. In the topographic surface, the “water” flooding process will stop, when the level of water has reached the highest peak, which leads to splitting the image into particular regions. The image gradient maps or Euclidean distance transform can be used with watershed segmentation to process the image (Chakkaravarthy and Chandrasekar, 2019)

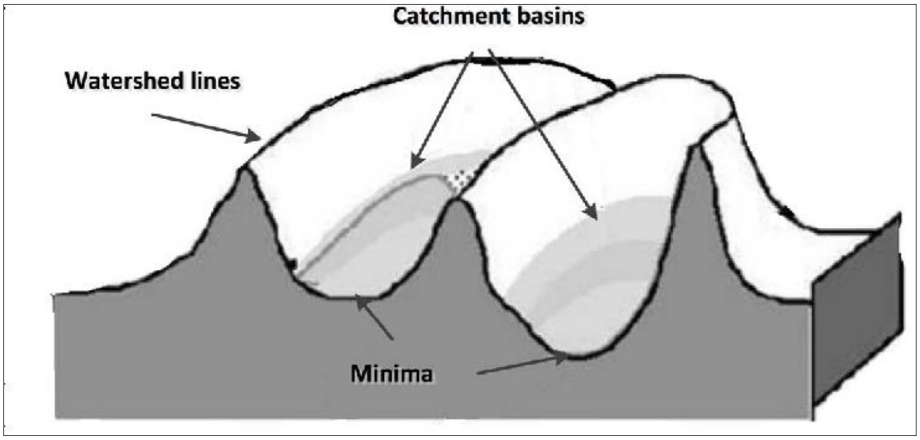


Figure 2.12 Watershed transform with watershed lines and catchment basins (Bai, Li and Wang, 2017)

2.4.2.3 Colour Segmentation

Colour-based segmentation is the process that segments or subdivides an image into regions depending on the colour-texture pattern distribution in that image (Russ and Neal, 2016). This process is about finding homogenous colour regions, and those not similar to the segmented region. Biological image segmentation should be accurate and robust, and subdivide an image to target objects and non-objects for further quantification.

Colour segmentation is different from one colour space to another as there are many colour spaces representing colour in image. Example of colour spaces are: RGB that includes three channels red, green and blue; CIELAB represents lightness, where A and B are two colour dimensions in this colour space. There are many other colour spaces, however the above are widely used in image analysis (Ford and Roberts, 1998).

Biological image segmentation is a difficult task because images are rich in colour and texture. Colour quantization facilitates the segmentation process, using a criterion for extraction of a few representative colours, where neighbouring regions can be differentiated in the image (Russ and Neal, 2016).

2.4.2.4 Mathematical Morphology

Morphological techniques can be used in the pre-processing and post-processing of geometrical structure in an image. Techniques can be applied to binary images and greyscale images. Mathematical morphology transforms the image according to the size, shape and connectivity. Binary morphology techniques control how the shape fits or misses the exact shape of the object in the image by using different morphological operations that help image analysis before and after segmentation (Gonzalez, Woods and Eddins, 2004; Fisher *et al.*, 2005; Meijering and van Cappellen, 2006).

- **Erosion and Dilation**

In the binary image, there are two common operators of mathematical morphology that are typically applied. The boundaries of foreground part will erode in the erosion operator. While the boundaries of foreground area will enlarge in the dilation operator.

- **Closing and Opening**

The processing by using close is same to the dilation operator which it reduce the background and expand the boundaries of foreground object in the same region. The closing operator can be utilized with two operations the erosion and dilation.

The closing operator in the similar shape background region doesn't affect to the structuring element. But the remaining areas are changed. In a simple define for the close operator is dilation then erosion, for both operations with the same structure element as shown in Figure 2.13.

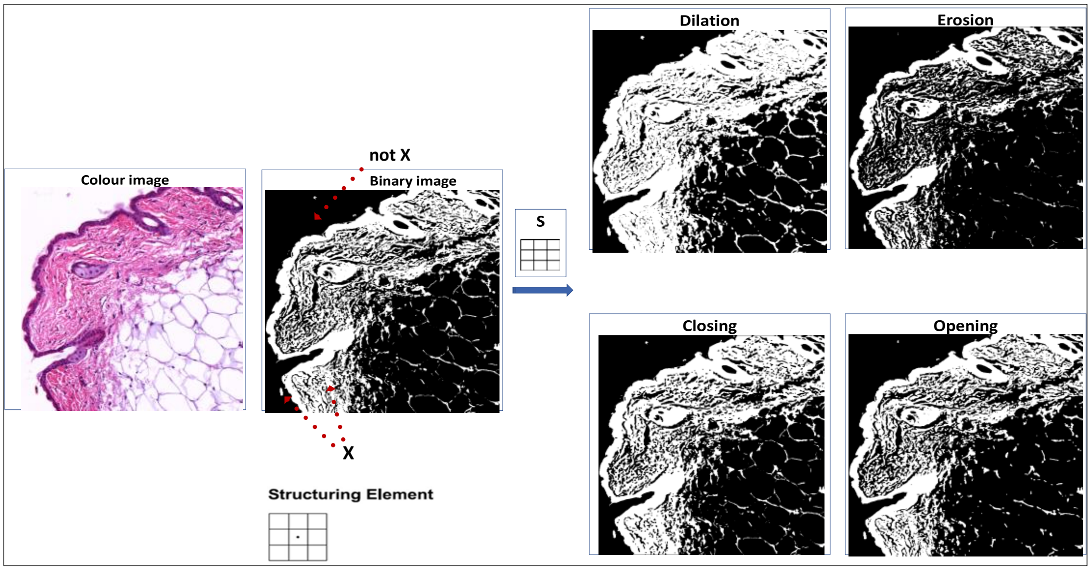


Figure 2.13 Principles of binary morphological erosion, dilation, closing and opening operators

2.5 Comparing Manual versus Automatic Measurements

To examine the correlation and the closeness between the automatic and manual measures, three different metrics are used.

2.5.1 Regression line

Describe the relationship between the actual and prediction measures is through fitting a regression line. We can use the R^2 (coefficient of determination) statistical measure, which can tell how the data is close to the fitted regression line to evaluate the correlation between the actual and predicted measurements (Frost, 2013). The R^2 can be calculated using the following formula

$$R^2 = \left(n(\sum xy) - (\sum x)(\sum y) \Big/ \sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]} \right)^2 \quad \textbf{\textit{(eq2.17)}}$$

Where x and y are two variables and n is the number of observations.

Roughly speaking, the higher the R^2 the better the fit. However, even the R^2 can explain the relationship between the two variables, it cannot give any information about the bias of the predicted variable. The R^2 is used as initial evaluation results in my thesis, such as in Chapter 5,6,7 to compare our automatic method and ground truth. Due to R^2 method does not give any information about the bias of the predicted variable, I used Angle of Regression Line (ARL) method to confirm out results.

To address the above problem, the Angle of Regression Line (ARL) is used. ARL gives beneficial information about the nature of relationship between the two variables. Usually, around 45° ARL indicates a good correlation of the predicted values to the actual ones. Smaller or larger than 45° ARL shows the bias of the estimated values.

However, the R^2 measure is good to show how two variables are related but not how much their values are close or agreed (Bland and Altman, 2010). For this reason, we have decided to use the Bland Altman analysis method as explained in next section for the closeness between different measurements.

2.5.2 Bland-Altman analysis

To further confirm our results The Bland-Altman (BA) analysis (Bland and Altman, 2010) is utilized to estimate the agreement between the results of the manual and automatic methods as I used in Chapter 5,6,7. The BA analysis is done as follow:

1. Calculate the mean and differences between the manual and automatic measurements.

2. Plot the difference between the measurements against their mean.

Calculating the bias (mean of differences ($\bar{d}$)) and the standard deviation (σ) of the differences between the two sets of measurements.

3. Calculate the lower and upper Limit of Agreement (LoA) as follow:
   - Upper LoA = $\bar{d} + 1.96\sigma$ ***(eq2.18)***
   - Lower LoA = $\bar{d} - 1.96\sigma$ ***(eq2.19)***

For the normally distributed differences, 95% of differences lie between the upper and lower LoAs. In general, when the automatic measurements are close to the manual ones, the bias, upper and lower LoAs should be close to the zero.

2.5.3 Mean absolute error

To predict error between our automatic result and ground truth. The Mean Absolute Error (MAE) is used as a metric to estimate the agreement between prediction and actual values as I used in Chapter 5,6,7. MAE for an n observations values could be given by the following formula.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|$$ (Chai and Draxler, 2014) ***(eq2.20)***

Where f_i is the prediction value and y_i is the ground truth value.

All the above mentioned statistical measures will be used to evaluate the differences between the ground truth measurements and the automatically estimated measurements for the proposed methods.

2.6 Object Classification Based On Machine Learning

Machine learning is increasingly becoming one of the essential fields of general practice and activity within computer science (Weir *et al.*, 2012)**.** One of the effect applications of various machine learning tools is in image analysis and segmentation. Using different algorithms and features extracted from the image, a model could be built to classify a specific objects such as nuclei in different layers based on extracted nuclei features by appropriate training an effective classifier for the recognition. In our research we use the support vector machine, which is a type of non-probabilistic binary classifier, which firstly presented by (Cortes and Vapnik, 1995) and since then have been used in many machine learning applications. It has been shown that SVM outperforms other classifiers such as Random Forest for nucleic segmentation in different layer of mice skin (Wen *et al.*, 2017). The classification of objects in standard SVM is based on decision boundary (hyperplane). Positive class objects are lying on one side of the line while the negative class objects are lying on the other side, which deems in (Figure 2.14). Out of many possible hyperplanes that divide the two classes, there is only one that maximise the margin. Indicated distance between the hyperplane area and the nearest site points on either part is recognized as a margin, which deems in (Figure 2.14). To decide a classification, the SVM is depending on the closest features to margin, which they are recognized as support vectors (filled circles and square which displayed in Figure 2.14) (Weir *et al.*, 2012)**.** The class of an example is determined by which part of the hyperplane the example appears.

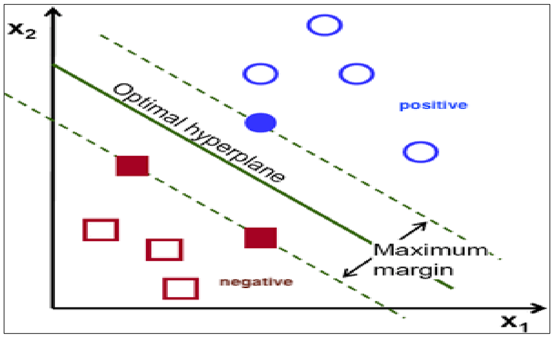


Figure 2.14 SVM classifier (Li *et al.*, 2013)

2.7 Existing Biological Image Analysis Tools

There are many existing tools for digital image processing and analysis, such as open-access software.

2.7.1 Free Open-source Software

ImageJ is a freely available tool used widely in biomedical image analysis. ImageJ is java-based open-source platform that has a list of plug-ins that make it helpful in a range of fields in science and engineering, including medical imaging and microscopy. Figure 2.15 a illustrates the user interface of this program. One important feature in ImageJ is that it supports a variety of standard image formats, including 48-bit colour composite images. LOCI is a bundle of plug-ins that was developed to open many different image formats from bioscience, (www.loci.wisc.edu). ImageJ has a number of useful tools for intensity processing such as image filtering (as illustrated in Figure 2.15 b and c), background subtraction and histogram manipulation, as well as a number of automated segmentation techniques such as Otsu thresholding, mixture modelling, maximum entropy, colour-based thresholding and k-means clustering (Collins, 2007; Baecker, 2010).

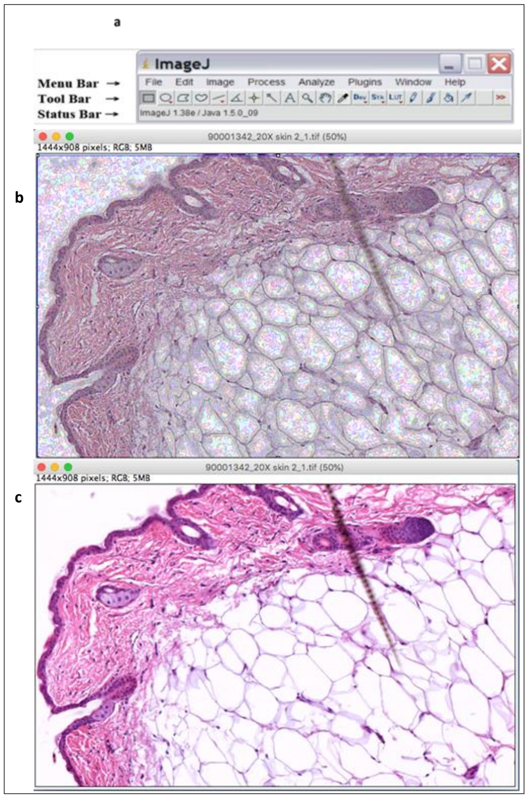


Figure 2.15 Noise reduction using ImageJ. a) The ImageJ window. b) An example of an image containing noise before using mean filters, c) the image filtered with a mean filter. Taken from (Baecker, 2010)

2.8 Summary

We reviewed the biological and technical backgrounds relevant to the research carried out in the thesis. The first part of this chapter develops a basic understanding of the genome projects. It also highlights the needs of MGP to sort of efficient and effective automatic image annotations to deal with information extraction from huge number of generated images.

The image segmentation techniques will help to segment the layers in the microscopic image, which is the first stage before do extracting the interesting objects, such as hair follicles. In addition these techniques may play important roles in post-processing of any segmentation results before we can extract a beneficial information from it.

Chapter 3 : Framework and Methodology

Having reviewed the biological and the technical backgrounds of the whole project in Chapter 2, this chapter focuses on high-level presentation of the main computational ideas and procedures adopted in the rest of this thesis. The descriptions are given in this chapter will help in understanding the logic behind our research approach in solving different problems throughout this thesis. This chapter describes the general framework of the research to highlight the philosophy behind the proposed solutions. It then presents the data sets and ground truth, and highlights the difficulties and the degree of complexity associated with ground truth. Different experimental protocols and evaluation methods used in this thesis are also discussed in this chapter.

The rest of this chapter is organised as follows. Section 3.1 gives an overview of the main steps followed in our proposed methods. In Section 3.2, an explanation of the data used is given. Section 3.3 highlights the ground truth available for our research. The experimental protocols for each problem are discussed in Section 3.4. Finally, a summary of the chapter is given in Section 3.5.

3.1 Research Framework

The main goal of this research is to produce fully automated solutions of high throughput data to annotate the microscopic H&E images of mice skin and to reveal/extract useful information (image and genetic features) from such annotations. The block diagram in Figure 3.1 illustrates the summary and outlined of the main stages of the thesis:

1- Automatic segmentation. In this thesis, many automatic segmentation is proposed based on the ROI and I state them in the following points:

2- Segmentation improvement. Several improvement on the segmentation has been done based the ROI.

Besides the above fusion scheme for improving layer segmentation, we may need to improve nuclei segmentation. First, the problem of nuclei overlapping in the epidermis layer is the main reason to improve the segmentation, by proposing automatic separation method to separate the overlap nuclei in the epidermis as ROI to obtain accurate number of nuclei. Secondly, not all segmented objects in the dermis are nuclei. The improvement to segment nuclei accurately in this layer is done by filtering the nuclei objects only to count them later depending on known biological fact about them.

3- Measuring and counting. When the final improvement of the segmentation is accomplished, nuclei counting in different layers, then different performance metrics are calculated from the orientation of the hair follicles, and measuring the curvature of the epidermis layer.

4- Applying experimental protocols. Depending on the annotated measurements taken from different objects in the skin layers, different protocols are applied to evaluate the goodness of the automatic annotation results.

5- Identifying cases of skin structure abnormalities, which leads to defining a list of candidate genes associated with the abnormalities.

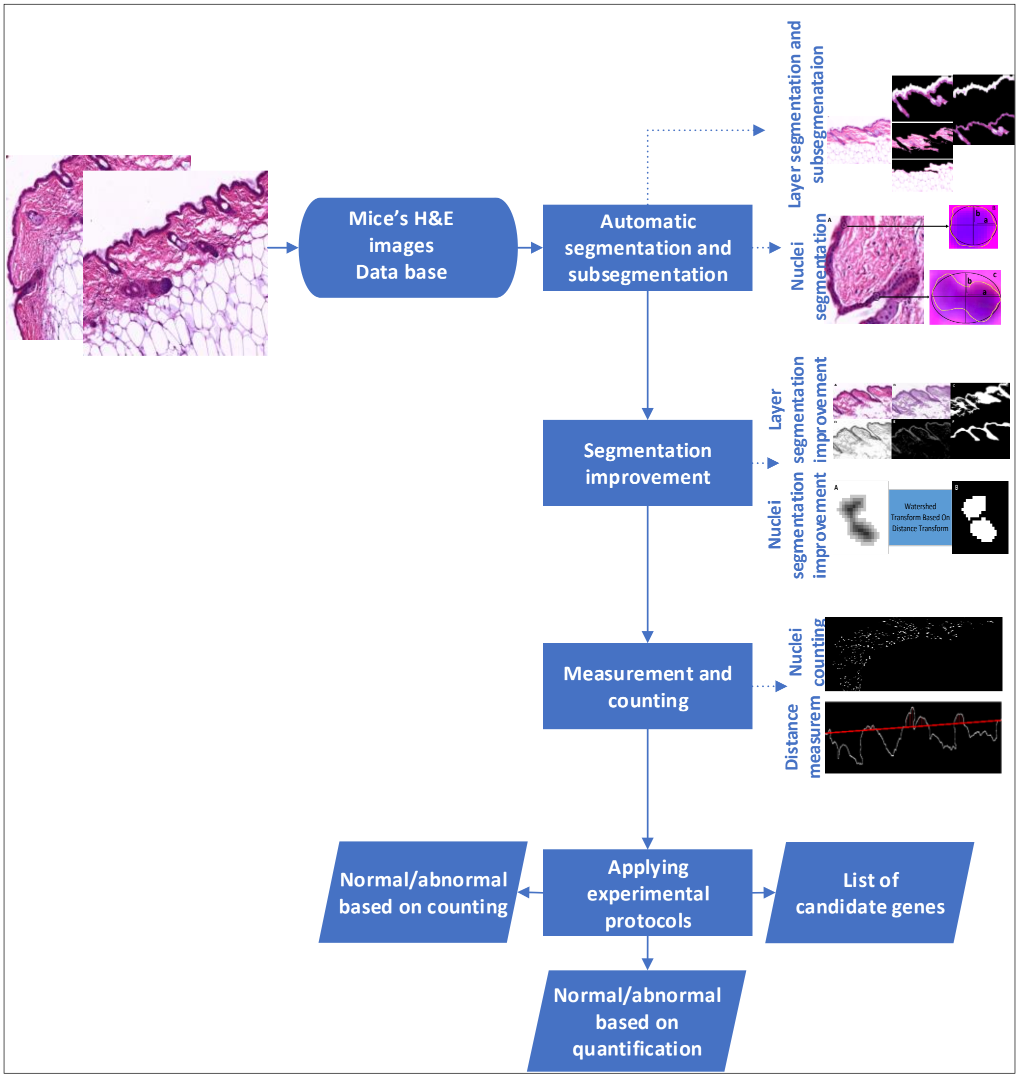


Figure 3.1 Research framework

3.2 Data sets

The data set used in this research includes H&E images of over 500 mice, which are including 7000 images. All images are colour images and their sizes are 1444*908 pixels. The data set is further divided into two sets. The first set includes 500 images of WT mice, providing the normal images of mouse skin. The second set contains 6,500 images of mutant mice obtained from 300 mutant lines (gene-knockout) where each mutant line has 14 images (7 males and 7 females). The data sets are provided by the WTSI, Sanger MGP (Liakath-Ali *et al.*, 2014; Takahashi *et al.*, 2019) The data publicly available for research use and hence we follow terms and conditions about data use as shown in Figure 3.2.

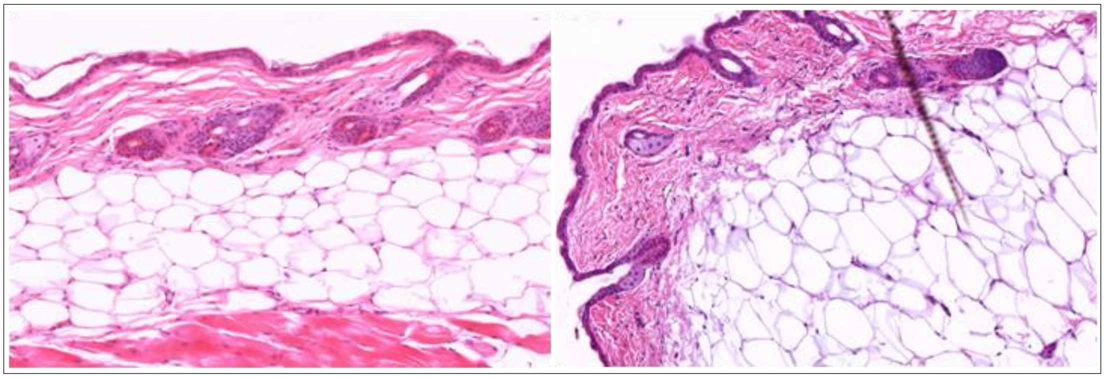


Figure 3.2 The two mouse H&E images views selected for the research

3.3 Ground Truth

In addition to the mice H&E image dataset, a ground truth file that contains the manual annotation for images was also provided by the Sanger Institute. The ground truth file contains a description for every mouse that determines its sex, barcode, genotype and also the genetic backgrounds. There are three types of genotype, WT (+/+) if no mutation in the gene, heterozygous (-/+) when the gene knockout is on one copy of the gene, if the gene knockout applied for both copies of the gene, this is called homozygous genotype (-/-). The file also contains a description for the full phenotype that results from a different gene knockout.

Unfortunately, the available ground truth file does not include all the needed information for our research. For the nuclei counting in different layers, there are no recorded measurements for the mice in terms of numbers, shape of the nuclei and also hair follicles orientation and the curvature in the epidermis layer have the same issue.

3.4 Experimental Protocols

In this thesis, there are different natures for the problems addressed, a number of experimental techniques and evaluation metrics are used to fit the requirements of the research. In the next subsections, descriptions about the evaluation techniques are given.

3.4.1 Identifying Genetic Markers using Reference Range (RR)

To robustly and reliably identify the genes that are associated with changes in a nuclei numbers, orientation of the hair follicles and the curvature, for the mice mutant group, the RR methodology based on percentiles is followed (White, Gerdin, Karp, Ryder, Buljan, Bussell, Salisbury, Clare, Ingham, Podrini, Houghton, *et al.*, 2013) to establish the normal range of WT variations. The RR is applied separately on the accumulated WT. The RR is established from the measurements obtained from applying the automated method on the 500 WT mice (normal) images. The Lower Bound (LB) and the Upper Bound (UB) of the 95% confidence interval of the WT measurements are set at the 2.5 and 97.5 percentile values respectively (White, Gerdin, Karp, Ryder, Buljan, Bussell, Salisbury, Clare, Ingham, Podrini, Houghton, *et al.*, 2013)**.**

To decide if a given mutant data is hit (i.e. it is outside the RR, which means the genes may be linked to the abnormalities), we calculate the so-called hit ratio of that mutant line. The hit ratio represents the percentage of images that falls outside the RR out of the total number of images in the mutant line (i.e. how many animals out of 7 per mutant line?). For example, the hit rate of a mutant line is 100% if all the 7 images of that line fall outside the RR. In this research, we set the hit rate threshold at 60% i.e. to consider a mutant line as a true hit, at least 5 out of 7 mice must fall outside the RR (White, Gerdin, Karp, Ryder, Buljan, Bussell, Salisbury, Clare, Ingham, Podrini, Houghton, *et al.*, 2013)**.**

3.5 Summary

In this chapter, a brief, and yet informative, description of the research framework of automatic annotation of mouse H&E images was presented. The framework includes the general ideas and thoughts to be applied in the following chapters. We also described the data sets and the ground truth that our proposed solutions will be tested upon. Finally, we described the main experimental protocols that will be used to evaluate the performance of the proposed methods solutions. In the next chapter, the first part of our work, which includes segmenting the main layers of mouse skin of H&E images, will be presented.

Chapter 4 : Layers Segmentation of Mice Skin

This chapter presents an automated high-throughput solution to segmenting mice skin layers using microscopic images. Such segmentation can be considered as a first stage to quantify cutaneous features that are known to be linked to certain skin diseases. For example, abnormalities in the nuclei in different layers (Feng *et al.*, 2019) . Furthermore, the existence of large datasets of skin images of gene knockout mice has made it possible to detect the genetic basis of some phenotypes in different dermal layers. However, using traditional manual methods to extract features related to a specific phenotype is time-consuming and could be subject to large inter-observer variabilities due to the subjective decisions made by different manual annotators. Therefore, there is a need to automated systems that are capable of generating reproducible results. The first stage of such systems would be segmenting the different layers of skin images.

This chapter has two main contributions:

1. Developing and test the performance of an adaptive skin multi-layers segmentation based on complexity of texture parameters. We shall illustrate that efficient methods such as color deconvolution are effective in segmenting images with simple texture whereas more sophisticated methods such as fuzzy C-mean clustering are needed to reliably segment images with complex texture. For the adaptive texture-complexity measure, we propose a simple model based on the number of image histogram peaks.
2. Proposing a hybrid system that combines the efficiency of colour deconvolution with the robustness of fuzzy C-mean clustering is proposed to segment skin layers from microscopic images.

The rest of the chapter is organised as follows. In Section 4.1, we talk about skin histopathological images of the WTSI Biobank datasets. In Section 4.2 describes the details of our proposed adaptive solution. In Section 4.3 we present experimental results achieved by performing colour deconvolution and fuzzy set clustering methods. The final section 4.4 concludes the chapter.

4.1 Introduction

The previous chapters illustrate that the WTSI (Wellcome Trust Sanger Institute) Biobank maximises the value of data on knockout mice for credible investigation of the genetic basis of some complex diseases via the development of high-throughput analysis (see Section 2.2 for details). The skin tissue Biobank is a resource of thousands of samples from mice in the pipeline that can be used to follow up any novel findings from primary and secondary phenotyping (H. *et al.*, 2016).

Mammalian skin in general is comprised of three main layers, the outer epidermis, the dermis, and the deeper hypodermis, as shown in Figure 4.1 (Goldman, 2008).

Disorder in skin layers due some diseases, such as skin cancer (Feng *et al.*, 2019), is the main motivation for distinguishing between normal and abnormal skin layers. Hence, the need to segment the layers of the skin individually as a first step in characterising the objects associated with the different layers.

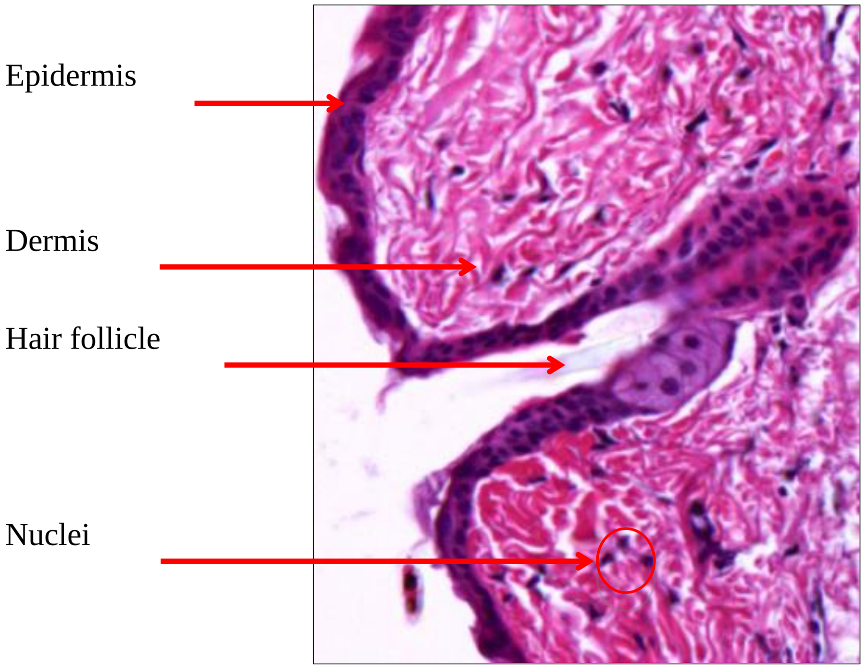


Figure 4.1 Microscopy H&E image of mouse skin (Goldman, 2008)

Accurate segmentation of skin layers is an important stage for the automatic measurements of relevant objects in each layer such as quantifying changes in the nuclei of dermis layer and hair follicle in epidermis layer as shown in Figure (4.1), and curvature quantification. Therefore, this chapter is mainly focussed on segment the three main layers in mice skin.

It can be argued that the quality of the layers segmentation has a direct impact on the quality of the measurements/features to be extracted from each layer. For example, if the segmentation does not produce a complete layer, it leads to inaccurate measure of the objects in the layer, such as incomplete epidermis layer in accurate counting of nuclei in this layer as shown in Figure 4.2.

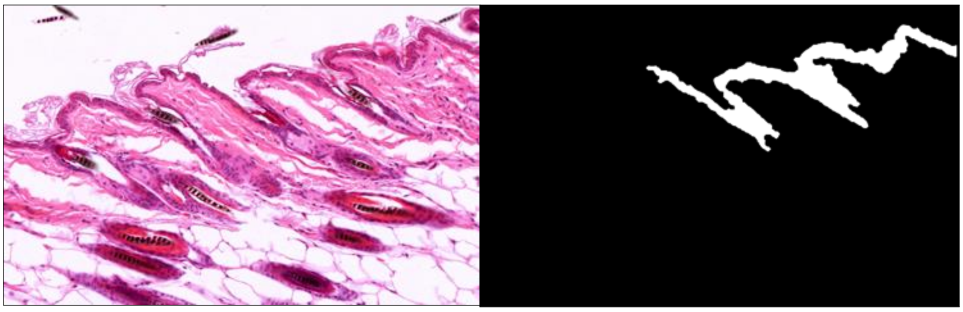


Figure 4.2 Incomplete layer segmentation. Microscopic H&E image (left image). Incomplete epidermis layer segmentation (right image)

Overlapping between the skin layers is another challenge to segmenting the layers. Figure (4.1) illustrates that layers of similar colour intensities are intertwined with each other. These issues make it difficult even for domain experts to precisely locate the borders of the layers. This means that manual measurement of skin layers will eventually encounter inter-observer and intra-observer variations if it was done manually.

Staining used in H&E images is known to be error prone, resulting in incorrect segmentation, particularly incomplete segmentation of layers or segmenting two layers as one (see Figure (4.3)).

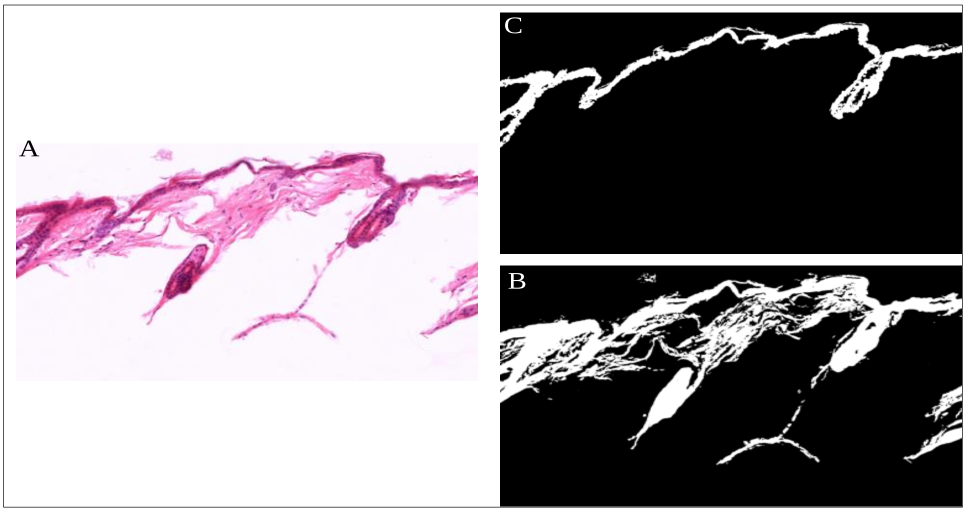


Figure 4.3 H&E stain error. A) Original image. B) Incorrect segmentation for two layers as on layer that are epidermis and dermis. C) Correct segmentation

An automated computer-based solution has to overcome the challenges highlighted above (overlapping and staining errors).

4.2 Our Proposed Method

Herein, we describe the development of two new layer segmentation methods so that adaptively one or the other can be used to segment automatic partitioning of H&E stained mouse skin images into their different skin layers: the colour deconvolution technique and the fuzzy system. We shall also define the adaptation parameter in terms of the complexity level of texture present in an input image to decide which of the two method is to be applied.

Designing new methods is due to the fact, that our investigation require the ability to segment not only the main layers but also the sublayers of these main skin layers Our proposed method is divided into three parts.

1. Investigating the effectiveness of a simple and efficient colour deconvolution based scheme to segment the layers in mice skin images, and determining is strength as well as shortcomings and limitations.
2. Increasing the reliability of the segmentation by using a fuzzy C-mean clustering algorithm to address the shortcomings/limitations of the colour deconvolution.

Proposing a hybrid system that combines the efficiency of colour deconvolution with the robustness of fuzzy C-mean clustering is proposed to segment skin layers from microscopic images. To combine the two methods, we propose a simple model to assess the complexity of the skin images, based the number of peaks in their histograms as illustrated in Figure (4.4).

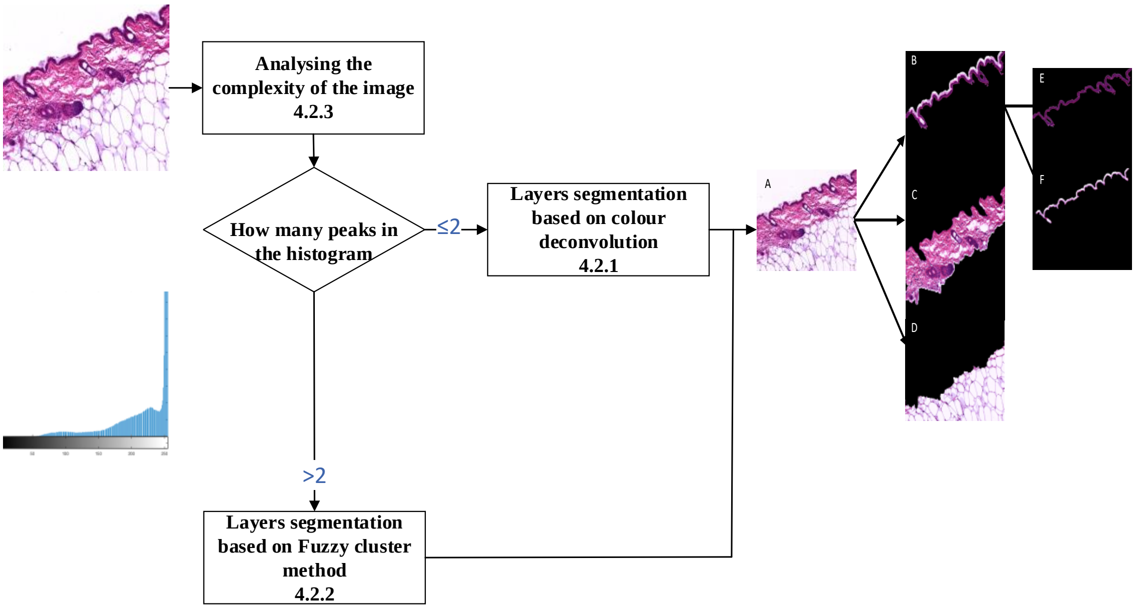


Figure 4.4 Research framework for layer segmentation

4.2.1Colour Deconvolution based Segmentation

Colour deconvolution is a method which separates overlapping spectra in cytological and histological (RGB) colour images into three new channels, compatible with (haematoxylin & eosin) H&E stain. This is considered one of the most successful segmentation methods related to the separation of different skin layers (Ruifrok, Johnston and others, 2001) as described in Chapter 2. However, it does not fit our essential requirement of ability to segments sublayers.

4.2.1.1 Epidermis layer segmentation

The epidermis layer consists of multiple layers, and for our work on skin related abnormalities we need to separate the epidermis into two layers, basal layer and cornified layer. This will enable our investigations on quantifying changes in the curvature of the cornified layer as clarified in (Chapter 7), as well as other disease-related skin abnormalities. The first step of the proposed segmentation algorithm is to segment the basal epidermis layer.

Segmentation of the basal layer of the epidermis

We applied the colour deconvolution transformation on the microscopic RGB images with H&E stain, which highlighted interesting layers such as epidermis and dermis. Whereas colour deconvolution is suitable with H&E(haematoxylin & eosin) stain image (Ruifrok, Johnston and others, 2001) (see chapter 2 for more details). The steps algorithm is stated in the following Pseudocode 4.1.

```
Step 1: Input RGB images.
Step 2: Apply colour deconvolution.
Step 3: Choose colour deconvolution one.
Step 4: Add image to itself twice.
Step 5: Binarisation images using Otsu's threshold.
Step 6: Dilation followed by an erosion using the disk structuring element with
        15  pixels (to segment basal layer accurately as a complete mask).
Step 7: Segment object with maximum perimeter.
```

Pseudocode 4.2 Steps for basal layer segmentation by using colour deconvolution

The colour deconvolution technique does highlight the H and E channel in separate images. As shown in Pseudocode 1 we used the colour deconvolution one which represents the H channel, which is highlighted in the epidermis layer.

The output of applying the colour deconvolution transformation as shown in (Figure 4.5B) is fed into a thresholding stage that separates the darker pixels (Basel layer pixels) from the background brighter pixels. The MATLAB Otsu Threshold function automatically implements clustering-based image thresholding (Chen *et al.*, 2012), results in reducing a grey-level image to a binary image that highlight significant objects/structures (Figure 4.5E). The algorithm works well on simple images that have bimodal histograms with two classes of pixels (foreground pixels and background pixels). In such cases, it computes the optimum threshold dividing the two classes so that their combined spread (intra-class variance) is minimal, and inter-class variance is maximal. A further pre-processing operation needs to be applied to segment the basal sub-layer, which has a maximum perimeter, and is determined by computing the boundaries of the labelled component and calculating distance between each adjoining pair of pixels around the border of the region. Similarly, the distance is computed for all the pixel positions for each object, and the object that has a maximum total distance, is considered a maximum perimeter. However, this process needs to be followed by the morphological operations of dilation and erosion in order to remove unwanted holes in the basal mask (Figure 4.5F).

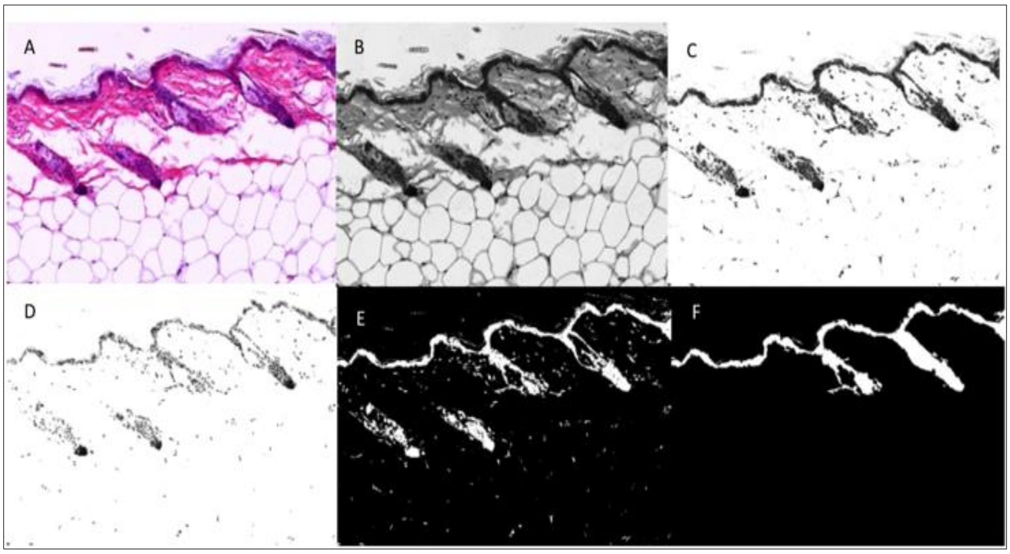


Figure 4.5 Automated image analysis methods to segment the basal layer of the epidermis. A) Original image. B) Image deconvolution colour one. C, D) Addition operation was used for image to itself twice. E) Binarisation using Otsu's threshold. F) Filtering out t

The entire process described above will be used to convert the H&E images into binary images, in order to demonstrate the region of interest for the basal layer of the epidermis.

Segmentation of the Cornified Layer of the Epidermis and Dermis

The cornified layer of the epidermis, also known as the stratum corneum, forms a barrier to save underlying tissue from infection, dehydration, chemicals and mechanical stress (Ovaere *et al.*, 2009). Inability to correctly maintain the skin barrier function due to the dysregulation of epidermal components can lead to skin disorders (Malik *et al.*, 2019). Due to the possibility that the change in the structure and function of the cornified layer could lead to disease, the development of an automated method to segment and measure the depth of this sub-layer provide potential benefits to histology-based research (Sanz-Gómez, Freije and Gandarillas, 2019). The segmentation algorithm for the isolation of the cornified layer is divided into two stages:

Stage one

In this stage, the cornified layer of the epidermis was segmented using deconvolution colour two, which highlights epidermis texture. The following block diagram show the algorithm steps for the stage one processing (Figure 4.6).

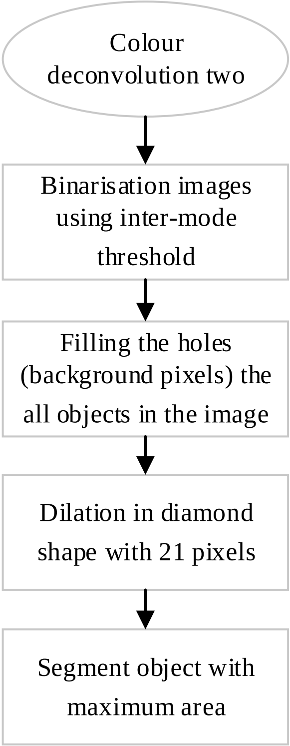


Figure 4.6 Algorithm steps for the stage one

As shown in Figure 4.6 we used Colour deconvolution two which represents the E channel, which highlights the dermis layer.

We used the inter-mode threshold to binaries the image because the inter-mode threshold is more effective on images with histograms that have extremely unequal peaks i.e. the inter-mode highlight the texture of dermis layer which make it easy to segment. The threshold found is considered as the minimum value in bimodal (Barreto, Tita and Orlandi, 2019). Next, a set of morphology operations were applied to avoid unwanted objects, for example filling the holes, to make a connected mask, the dilation in diamond shape 21 pixels as the size of dilation used to segment accurate complete mask. If dilation is more than 21 pixels, the algorithm will segment another layer with dermis layer such as fat cells layer. If dilation is less than 21 pixels, algorithm will not segment a complete mask of dermis layer i.e. the output segmentation will be part of the dermis mask. Finally, counting the number of white pixels (i.e. logical value=1) of each binary object and selecting the one with the maximum area, as shown in Figure 4.7

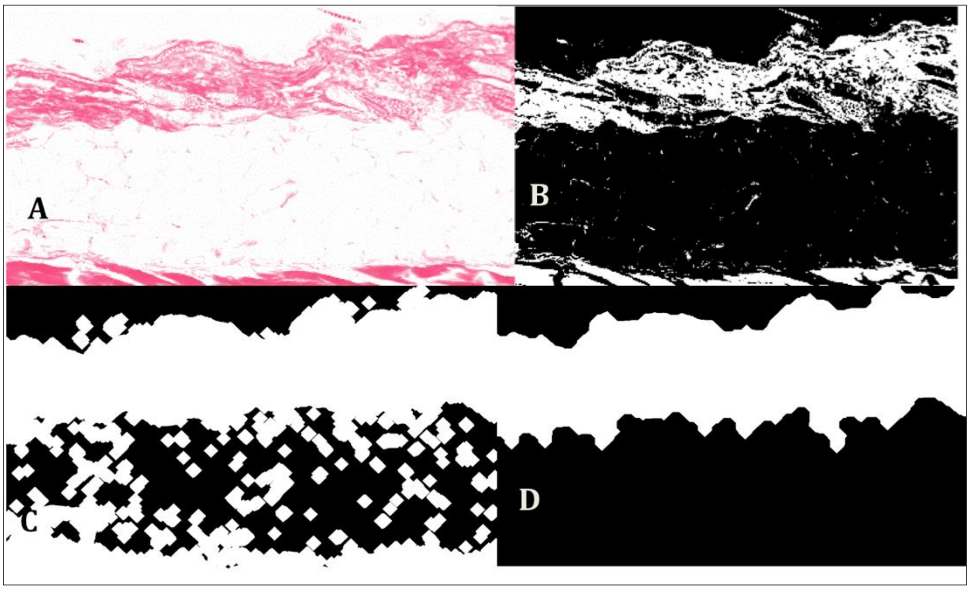


Figure 4.7 Automated image analysis methods to segment the mask one. A) Image deconvolution H&E colour two. B) Threshold using inter-modes method. C) Morphological operations. D) Select the object, which has a maximum area

Stage two

The Pseudocode 4.2 of the stage two is showing in the following box

```
Step 1: Logic invert of basal mask.
Step 2: Multiplications for each pixel between logic invert of basal mask
        with dermis and epidermis mask.
Step 3: Segment the object which has a minimum Y (cornified layer).
Step 4: Multiplication for each pixel with logic invert the epidermis layer
        to segment the dermis layer.
```

Pseudocode 4.2 Steps for segmenting dermis layer using colour deconvolution

In the second stage, the epidermis mask from stage one (Figure 4.8A) is multiplied by the logical invert of the basal layer mask (Figure 4.8B). Then the morphological operation was applied to segment the cornified layer (Figure 4.8D).

The dermis layer was then found by the multiplication operation for each pixel between the logical invert of the epidermis layer mask (Figure 4.8E).

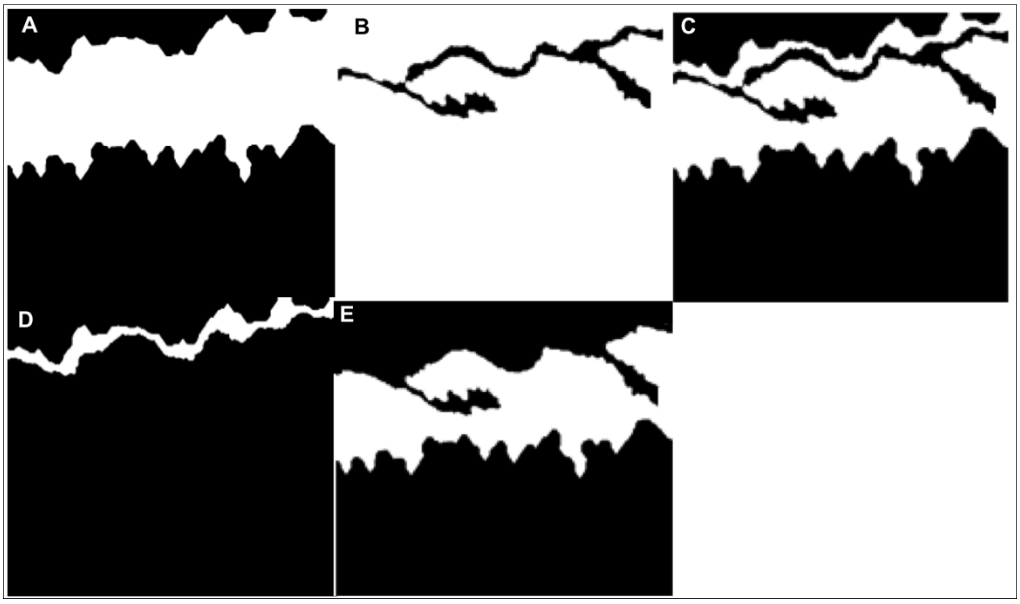


Figure 4.8 Automated image analysis methods to segment the cornified layer. A) Dermis and epidermis mask. B) Epidermis (basal layer) mask. C) Multiplications between A&B masks. D) The object, which has a minimum Ywhich is, Cornified layer with epidermis. E) Dermis

4.2.1.2 Segmentation of the adipose layer

The adipose layer is located beneath the dermis (Wang *et al.*, 2013). The following diagram is describe the steps to segment the fat cell layer (Figure 4.9).

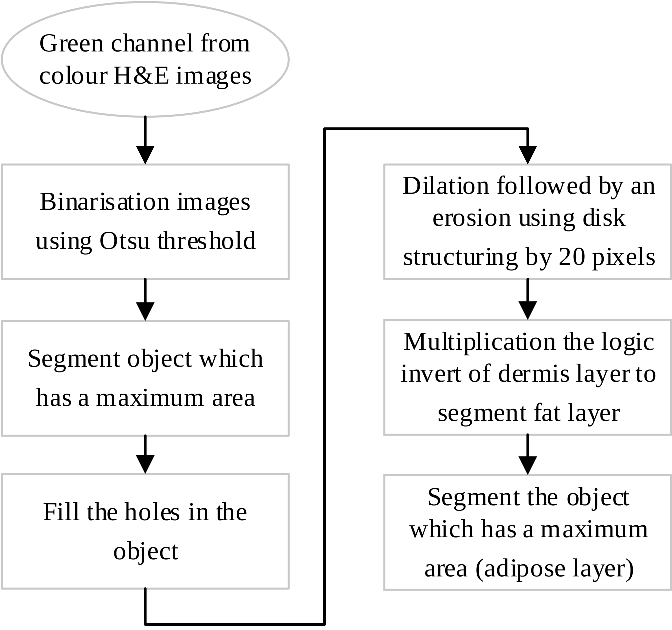


Figure 4.9 Steps to segment adipose layer

The green channel from the colour H&E image was selected (Figure4.10A), and the Otsu automatic threshold is applied (Figure 4.10B). Next, a morphological operation was performed by segmenting the object, which has the maximum area to avoid unwanted objects from the binary image (Figure 4.10C). Finally, a set of morphological operations was used to segment the fat cell layer or subcutaneous layer such as dilation 20 pixels, to segment accurate subcutaneous mask and avoid segment another layer with subcutaneous such as dermis layer. The images shown in Figure 4.10 (D and E).

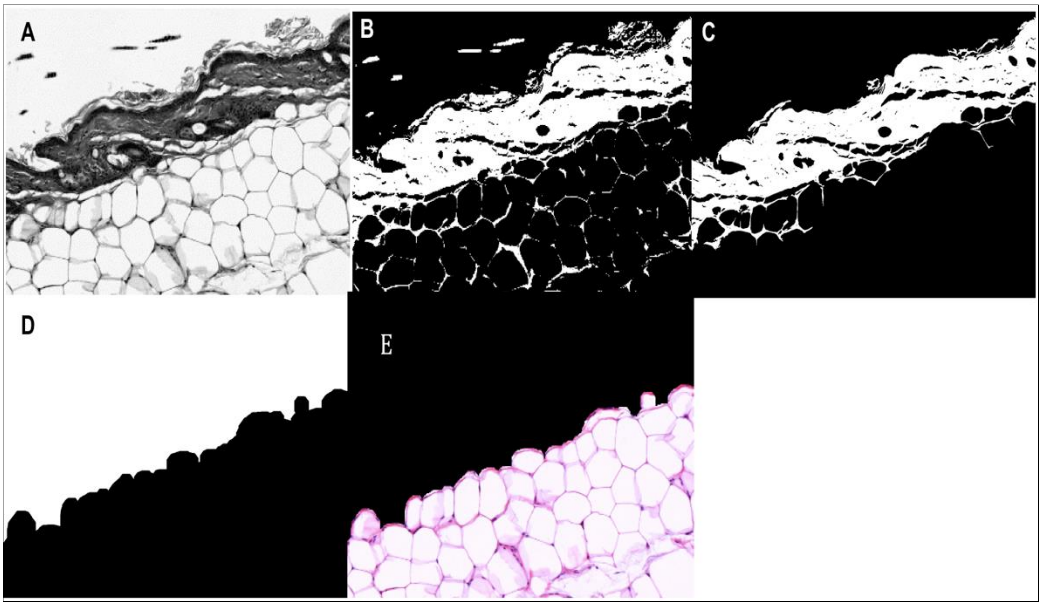


Figure 4.10 Automated image analysis methods to segment the adipose layer. A) Green channel image. B) Automatic threshold image. C) Selected the object, which has maximum area. D) Set of Morphological operations. E) Colour adipose image

4.2.1.3 Discussion

Colour deconvolution is one of the most common method used image processing function to isolate overlapping colour. It is suitable for H&E (haematoxylin & eosin) stained images (Bulten *et al.*, 2019). Two types of threshold were employed to segment the layers, Otsu and inter-modes thresholding. Otsu thresholding was used to segment the epidermis layer, because the Otsu thresholding highlights the epidermis rather than dermis layer. While the inter-modes thresholding was used to segment the dermis layer, because the inter-modes threshold highlights the dermis rather than epidermis layer. The colour deconvolution method, thresholding and segmentation allowed the development of an effective automatic layer segmentation system for H&E images in our proposed method.

We noticed that when the histogram of the H&E image is somehow complex i.e. there are more than two peaks in the histogram, the colour deconvolution method described above method does not work (around 30%) as illustrated in Figure 4.11B and C. Therefore, there is a need to find another method that is capable of segmenting complex images with more than two peaks in the histogram. However, our literature survey showed that the fuzzy c-mean method does exactly that, and hence we describe in the next section. However, the fuzzy c-mean method does not work well on simple images, which suggest the need for combine the colour deconvolution method with the fuzzy cluster to achieve good accuracy on all images but using an adaptive approach, whereby, depending on some texture complexity indicator one of the two algorithms is applied.

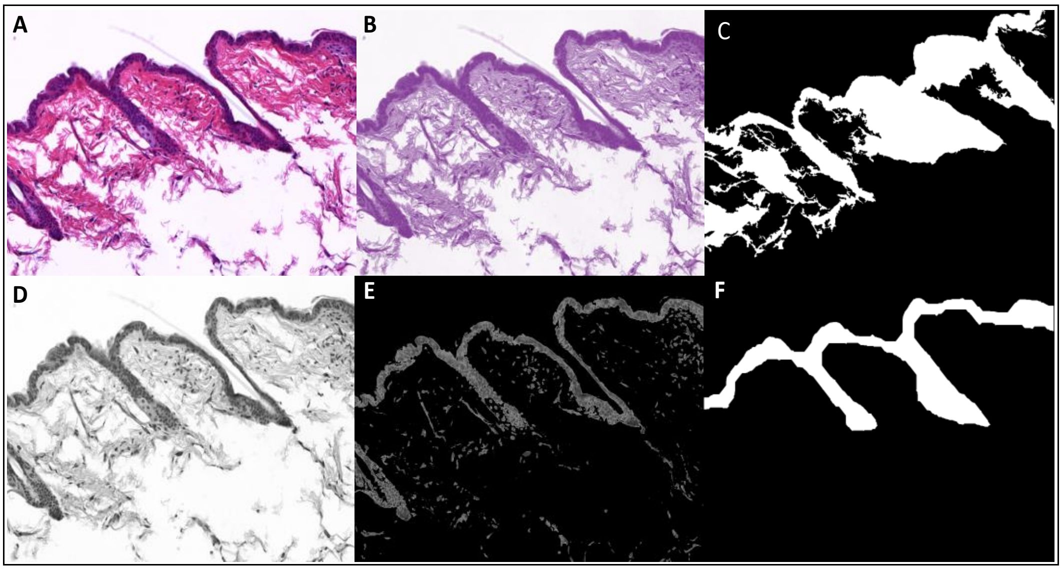


Figure 4.11 Method of segmentation A.) Original H&E image. B) Colour deconvolution H channel segmentation of epidermis layer. C) Image segmentation using Otsu threshold. D) Red channel of H&E image. E) Cluster segmentation based on fuzzy c-mean method. F) Epidermis layer

4.2.2 Layer Segmentation based on Fuzzy C-Mean Method

As explained above, the colour deconvolution does not work well on images that have complex texture, e.g. when there are overlaps, in terms of colour intensities, between the dermis and epidermis layers in skin images.

To address this issue, a more robust method to cluster the pixels is needed. The Fuzzy C-mean method (FCM), as described in Chapter 2 - Section 2.4, groups image pixels in particular clusters according to distance and membership. It was shown by (Zhang *et al.*, 2017) that the FCM is capable of segmenting the skin layer of human images correctly. Therefore, it is a natural choice for our investigation. It is possible, for example that the overlap areas of the dermis and epidermis layers appear in visible clusters. However, our version of the FCM algorithm works in steps: (1) segment the epidermis layer and sub-segmentation into two layers (the basal layer and cornified layer), (2) segment the dermis layer, and (3) finally segment the fat cell layer (adipose layer). The following is a block diagram for this algorithm (Figure 4.12).

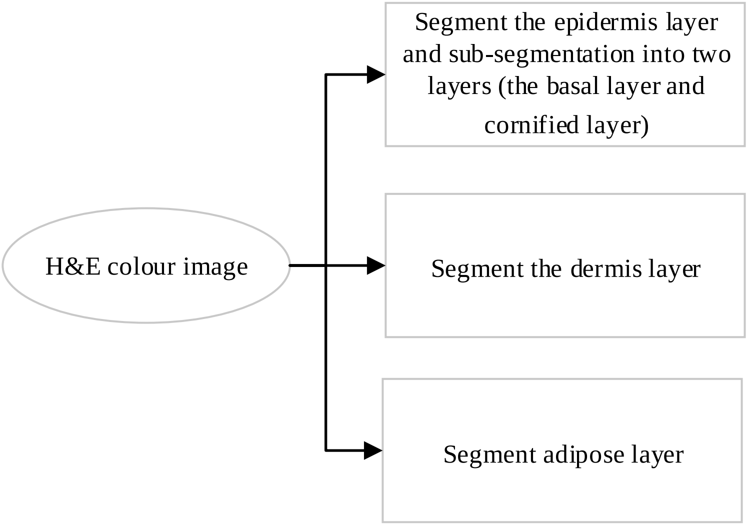


Figure 4.12 Block diagram shows three layers segmentation steps

4.2.2.1 Initial Segmentation with threshold level

The threshold level of three clusters of fuzzy c-means clustering have used as automatic segmentation method. The threshold level was found by determining the average between the maximum of the small peak and the minimum of middle peak (Xiong *et al.*, 2006; Chen, Li and Zhou, 2007). The binary output is 0 when the level of threshold was located between the small peak and the middle peak whereas the output is 1 when the level located between the middle peak and the large peak.

The three clusters resulting from the Fuzzy C-Mean method are illustrated in Figure 4.13. Cluster one highlights the epidermis layer and the nuclei in all layers. Cluster two shows a texture for all layers as seen in Figure 4.13D, and cluster three highlights the background of the image as shown in Figure 4.13C.

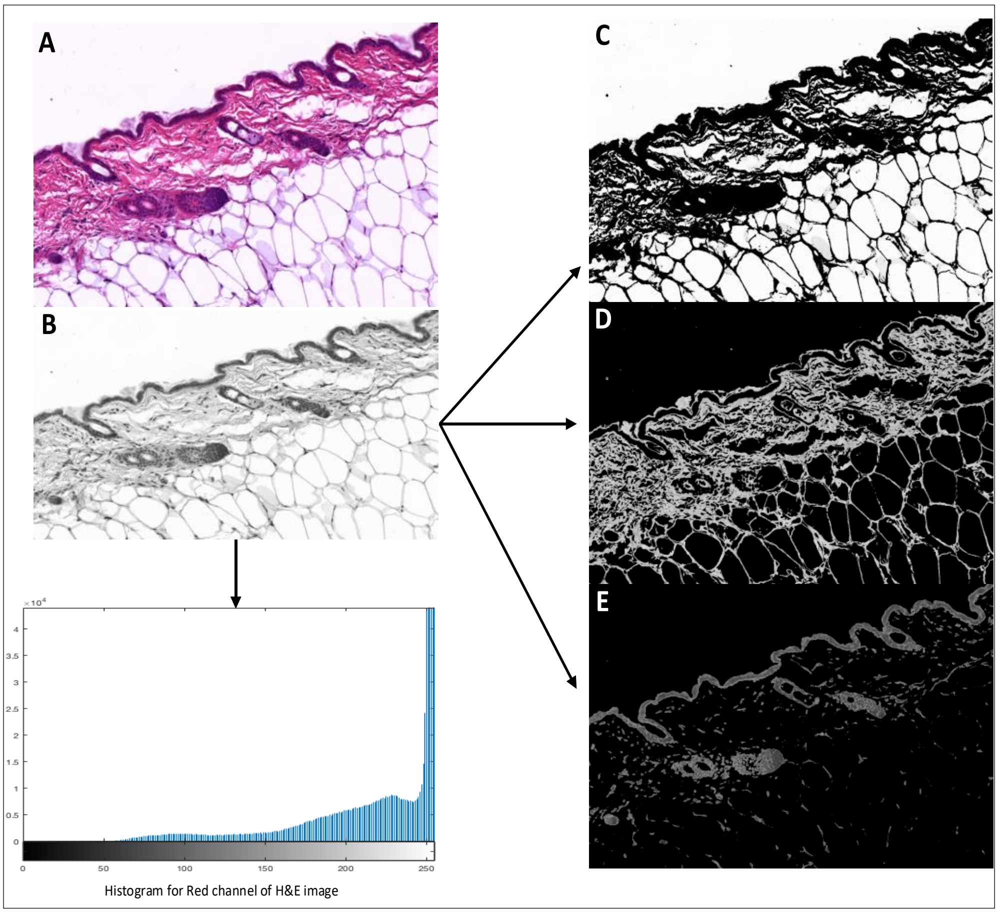


Figure 4.13 Outcome the FCM based on histogram. A) Original image. B) Red channel. C, D, E) Three clusters by FCM method based on the histogram of red channel of H&E stain image in the mouse skin with 20X magnification

We argue that the threshold level has managed to capture the area of interest, without losing important textural details of the skin layers. On the other hand, the Otsu's method (Du, Chen and Xi, 2019) split the histogram into two classes, the foreground and background, which minimises the variance inter-class, so that will result in losing parts of the area of interest.

4.2.2.2 Epidermis layer segmentation

The first stage of layer segmentation involve splitting the epidermis layer by sub-segmenting it into two layers, namely the basal layer and the cornified layer. This procedure was initiated by using the FCM method of the red channel based on the histogram of H&E image by contributing an automatic method to detect the peaks of the red channel image as a following Pseudocode 4.3.

```
Step 1: Input RGB images.
Step 2:  Determined red channel of H&E image.
Step 3:  Apply fuzzy c-mean method with three clusters.
Step 4:  Output three clusters.
```

Pseudocode 4.3 Steps for epidermis layer segmentation using FCM

It can be clearly observed that the red channel of the image consists of more than two peaks of representation for distribution values for the pixels in the image. Subsequently the FCM technique was applied on the red channel of the image, to construct three clusters of the image, as shown in Figure 4.13 (C, D, and E). Cluster C represents the background of the whole image, while cluster D shows the texture for skin layer segmentation, and cluster E displays clearly the epidermis layer, where we are interested. The reason for utilising three clusters instead of two clusters was because of the results of FCM for two clusters, which is shown in Figure 4.14 (B, C), where image B represents the cluster which consists of the texture for all skin layer, which makes it hard to separate the epidermis layer and sub-layer, because there were overlap between the dermis and epidermis layer with approximate similar colour intensity, that is why we did not use two clusters in the FCM method. As mentioned previously, we used three clusters, where the image E cluster in Figure 4.13 clearly shows the epidermis layer with the nuclei of dermis layer.

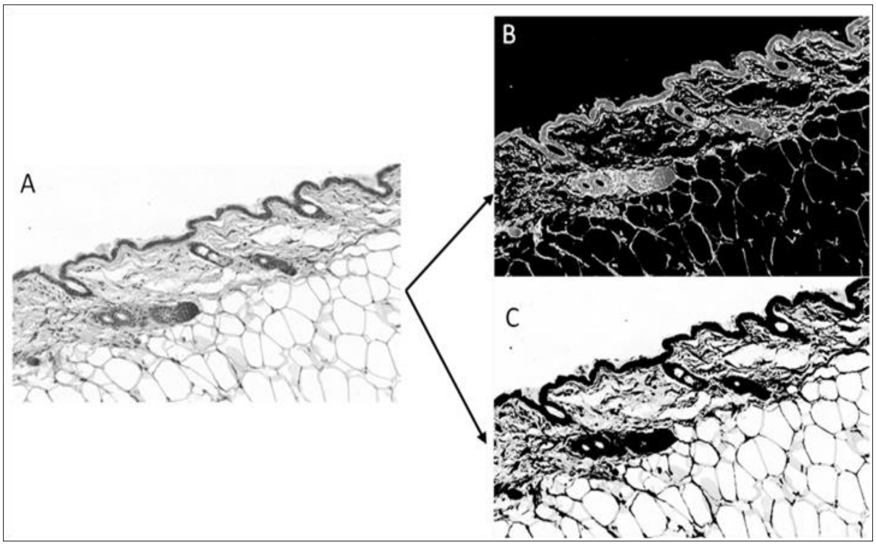


Figure 4.14 Result of FCM methods for two clusters based on red channel. A) Red channel. B, C) two cluster

Pre-processing phase of three clusters

There is a pre-processing phase of the three resultant fuzzy clusters, as illustrated in Figure 4.13(C, D and E). The E cluster which has the epidermis layer also contains the ROI (region of interest), and the cluster consists of the objects in greyscale colour as imaged contents and a black pixel as the background of the image. It is clear that the number of black pixels overshadows the number of colour pixels in image E. As a result, counting the number of black pixels involved a precise calculation technique to identify the best cluster for the epidermis layer segmentation. The number of black pixels represents the majority of the image, as evidenced by the overwhelming black colour on the image, so it can be inferred that the epidermis cluster E has the maximum number of black pixels (1196184) compared to the results of other clusters of the fuzzy method which were (980474) pixels in cluster D and (445646) pixels in cluster C. The following is a block diagram for this algorithm (Figure 4.15).

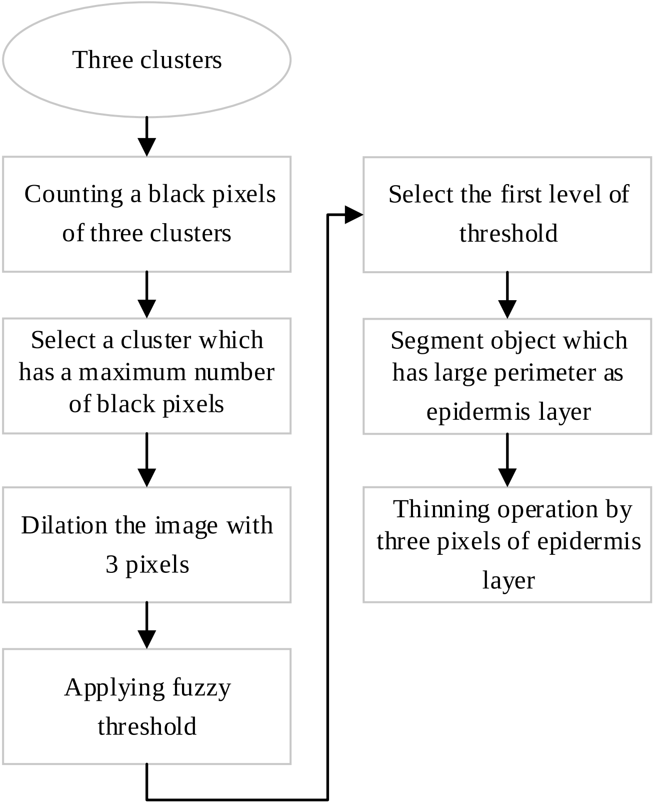


Figure 4.15 Block diagram show the steps of epidermis segmentation

Following this, a set of morphological processing was used such as a dilation operation by adding three pixels to the boundary of object, in order to highlight the epidermis layer of the selected greyscale image B in (Figure 4.16). The image is then binarized by applying the fuzzy threshold with three classes of clustering, i.e. the threshold based on three peaks in the histogram of images, and it will generate two levels of images based on three peaks, where the output of binary image was 0,and the level of threshold between the small peak and the middle peak (Mandal, Gupta and Kar, 2016) as illustrated in Figure 4.16C. The level between the middle peak and the large peak as shown in Figure 4.16D is not used because it segments the binary border of the objects in the image, while the level was selected based on the middle peak and the large peak, which consists of most of the pixels with a high value of greyscale, representing light colours. After thresholding, the object which has a large perimeter was selected, which represents the epidermis layer in Figure 4.16E. A set of morphology operations was applied, such as the thinning operation by three pixels, which keeps the curving of the binary mask, to avoid noise of the epidermis layer, as shown in Figure 4.16F.

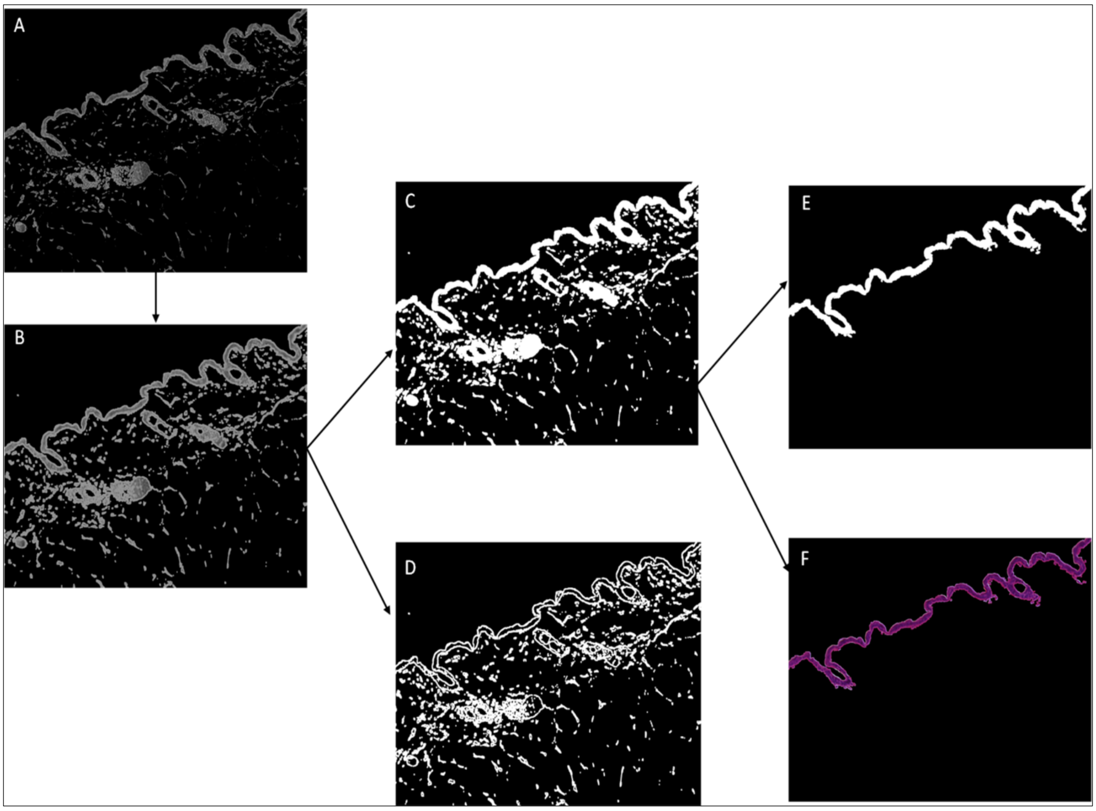


Figure 4.16 Epidermis layer segmentation. A) Cluster image, which has maximum numbers of zero pixels, which represent the epidermis layer cluster. B) Dilation the image with 3 pixels. C) Threshold based on FCM for three clusters, which produce two levels of threshold, the first level in image C, where the position of level in the histogram is between a small peak and a middle peak, the output binary image will be 0, which represents the core of the epidermis layer. D) The second level of threshold is based on the histogram of the image, where the level position is between the middle peak and the large peak, then the output binary image will be 1, which represents the boundary of the epidermis layer. E) Segment the epidermis layer by selecting the object, which has a larger perimeter in the image

Segmentation of Epidermis Sub-layers

The sub-segmentation of the cornified layer from the epidermis layer starts with the use of the dilation operation for the epidermis mask layer of 10 pixels with (disk) structure to cover exactly the epidermis layer, if dilation use more than 10 pixels, it will segment another layer with epidermis such as dermis. If dilation use less than 10 pixels, it will loss part of epidermis mask as shown in Figure 4.17(B). The aim is to make a mask cover of the sub-layer, located at the boundary part of the epidermis layer. A multiplication operation for each pixel was then applied between the epidermis layer after dilation and the invert of the epidermis layer before dilation to filter the cornified sub-layer by avoiding the epidermis layer, which can be seen as the output in image D of Figure 4.17. Following this, labelling of the objects in the image is carried out in order to select the object which has a minimum Y of the image for a final segmentation in Figure 4.17 (E, F).

This algorithm is summarized in the following pseudocode 4.4.

```
Step 1: Input binary mask of epidermis layer.
Step 2: Dilation operation for the epidermis mask layer of 10 pixels with
        (disk) structure.
Step 3: Inverting of epidermis layer mask before dilation.
Step 4: Multiplication operation between two masks (step 2 and step 3).
Step 5: Sub-segment the cornified layer (part of epidermis) by obtaining the
        object which has a minimum Y.
```

Pseudocode 4.4 Steps for sub-segmentation of epidermis layer

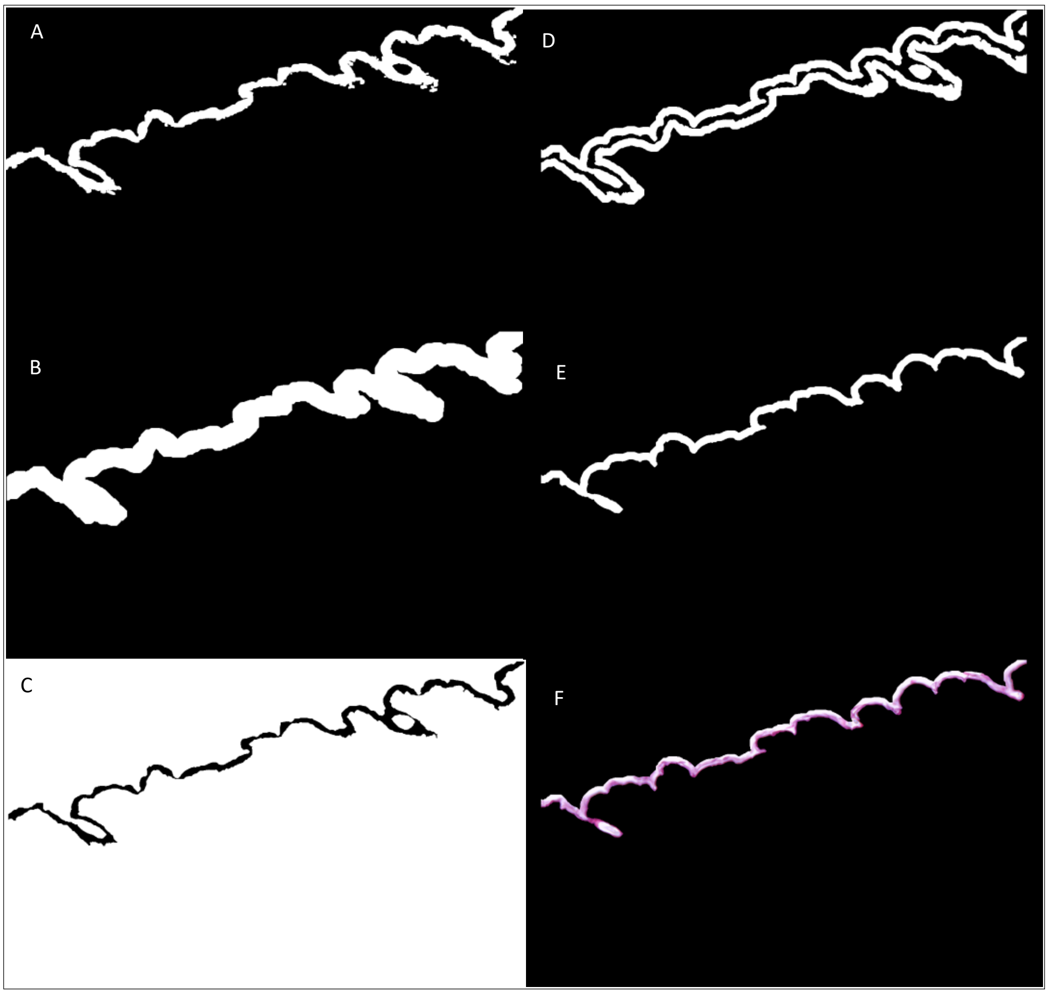


Figure 4.17 Sub-segmentation of epidermis layer. A) Binary mask of epidermis layer. B) Dilation of epidermis layer by 10 pixels. C) Inverting of epidermis layer mask before dilation. D) Multiplication operation between two masks (B, C). E) Sub-segment the cornified layer (part of epidermis layer) by obtaining the object which has a minimum Y. F) Colour mask of cornified layer, based on the original colour image

4.2.2.3 Dermis Layer Segmentation

The second threshold level of the FCM method was utilised in Figure 4.20D, which is located between a middle peak and large peak in the histogram of H&E colour image, where cluster three is used as displayed in Figure 4.19 (C), which highlights the dermis layer and other textures as black pixels in the image. The first threshold level was not used in image B in (Figure 4.19), because it is clear that in image B, the black pixel doesn’t cover the entire dermis layer, so it is not useful for the purpose of this study. The close gaps has been filled using the holes structure in image E (Figure 4.19), after which, the biggest perimeter object was segmented in image F (Figure 4.19). Following this, three pixels erosion were removed (by using erosion) from the boundary of the dermis mask, to avoid unwanted objects, which are not in the dermis layer, as shown in Figure 4.19 (G).

A filtering operation was applied to the dermis mask by using the epidermis mask to segment the dermis layer without the epidermis layer with 5 pixel dilation to cover dermis layer, as displayed in Figure 4.19 (H). Finally in image I, the large perimeter object was segmented to obtain a final dermis layer, as shown in Figure 4.19 (J).

The following is a block diagram for this algorithm (Figure 4.18).

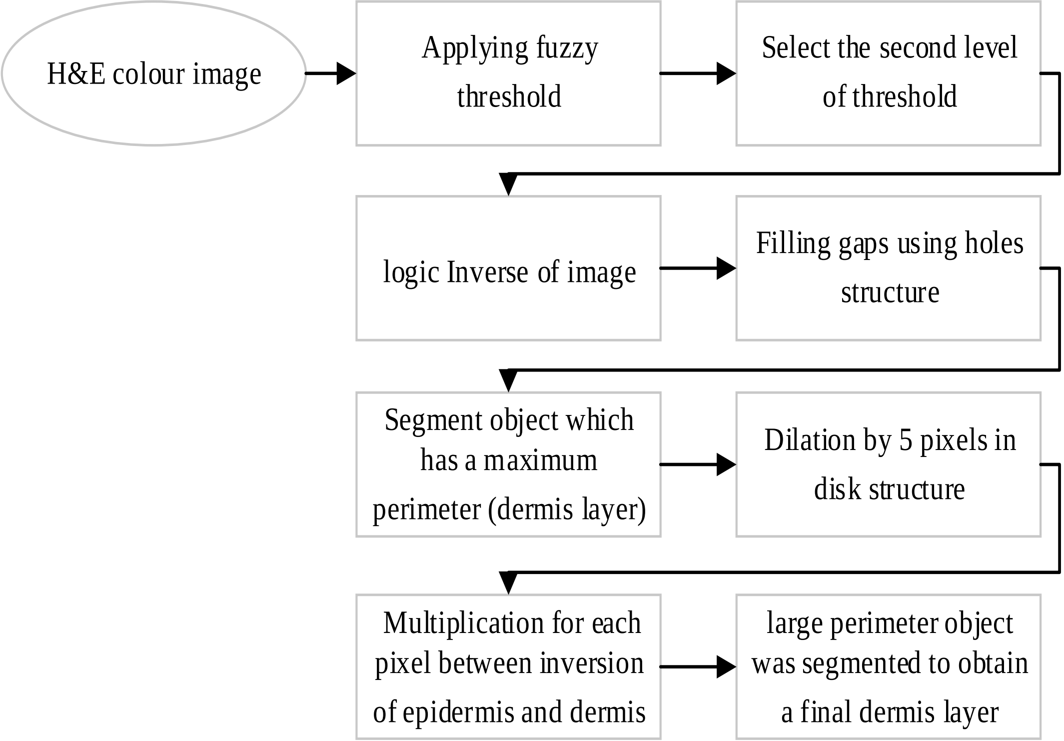


Figure 4.18 Block diagram show the steps of epidermis segmentation

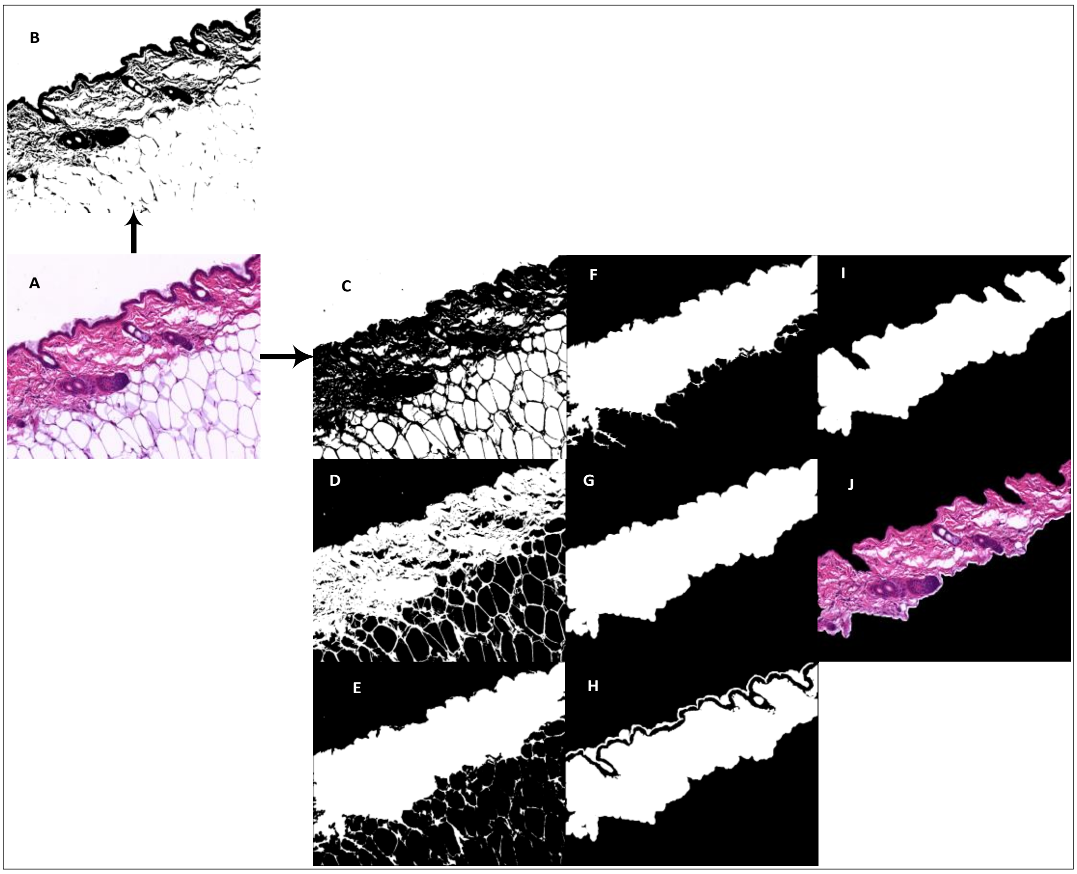


Figure 4.19 Dermis layer segmentation. A) H&E image B) Threshold one of FCM method, where it is located between the smallest peak and the middle peak, therefore it is not useful in this work because it doesn't cover the dermis layer. C) Threshold two of the FCM method based on the histogram, where it is located between the middle peak and the largest peak. D) Invert of image C to enable processing of the image in the next stages. E) Image after erosion by 3 pixels in disk structure. F) Larges perimeter objects have been selected. G) Dilation operation was applying based on 5 pixels in disk structure. H) Multiplication operation for each pixel between the inversion of epidermis mask and the dermis mask. I) Final binary mask of dermis layer. J) H&E dermis layer segmentation

4.2.2.4 Fat Cell Layer Segmentation

The fat cell layer is segmented by using the second threshold level of the FCM method as illustrated in image A of Figure 4.20, while the output of inverting image A was represented in image B, to provide a manipulated image to use in the image processing techniques. Next, a multiplication operation was conducted for each pixel in the image between the dermis layer mask, as shown in image C and the binary image B, and the result was represented in the image D, which consists of the fat cell layer with some noise objects that don't belong to the fat cell layer. Finally, dilation by 80 pixels (to cover all subcutaneous layer accurately) with the disk structure, and the noise was ignored by segmenting the object that has the largest perimeter in the image as displayed in image E and F of Figure 4.20 as a final segmentation. This algorithm is summarized in the following Pseudocode 4.5.

```
Step 1: Input RGB images.
Step 2: Apply fuzzy cluster threshold method.
Step 3: Select the second level of threshold.
Step 4: Invert the output image of step3
Step 5: Multiplication operation between dermis layer mask and the output
        mask from step4.
Step 6: Dilation operation by 80 pixels with the disk structure.
Step 7: Segmenting the object that has the largest perimeter in the image as
        finial fat cell layer segmentation.
```

Pseudocode 4.5 Steps for Fat cell layer segmentation

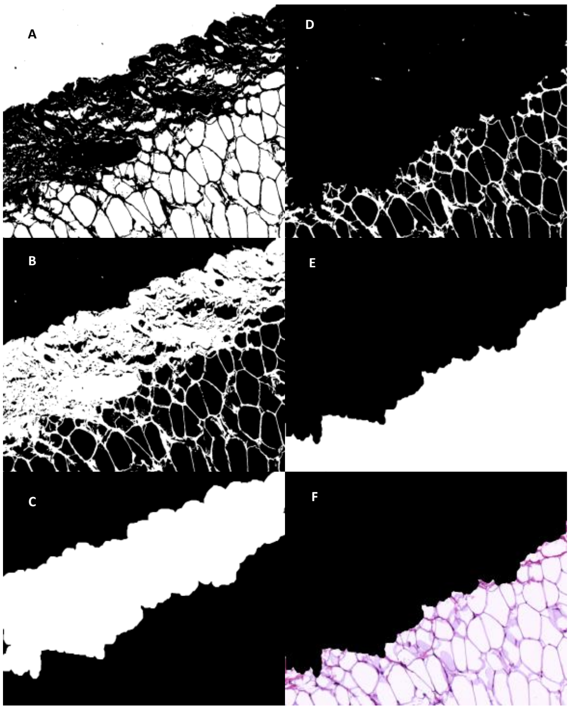


Figure 4.20 Fat cell layer segmentation. A) Threshold second level based on FCM. B) Inverting image A to enable the image processing method. C) Binary mask of dermis layer segmentation. D) Fat cell layer with noise does not belong to the fat cell layer. E) Segmentation of fat cell layer without noise. F) Fat cell layer with H&E stain

4.2.2.5 Discussion

The FCM method was a creative technique used to segment skin layers in the H&E, however, it has some shortcoming that results in wrong segmentation for some images such as (10 out of 100 images). For example, image B in Figure 4.21, illustrate incorrect segmentation, because the FCM seem to be constrained by the intensity change in the illumination of the pixels (Yambal and Gupta, 2013). Furthermore, the damage stain samples were fundamental issues of the segmentation methods, as shown in Figure 4.22 (A, B, C). Another issue of FCM is to related to the speed of the algorithm being too slow consuming more time than the colour deconvolution method, as used in Subsection 4.2.1. The time consumed for segmenting 10 randomly selected images, for two methods, is shown in Figure 4.23. Evidently, the time consumed in the FCM technique is more than that of the colour deconvolution method. The delay in implementation was likely due to the iteration steps with calculations, such as computing the membership for each pixel and the distance between the centroid and the cluster (Yambal and Gupta, 2013).

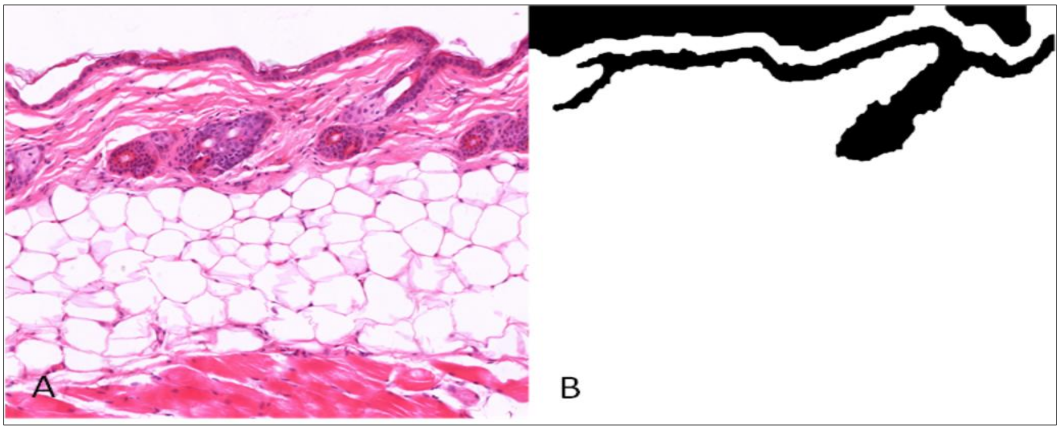


Figure 4.21 Incorrect segmentation.) A Original H&E image. B) Incorrect mask segmentation

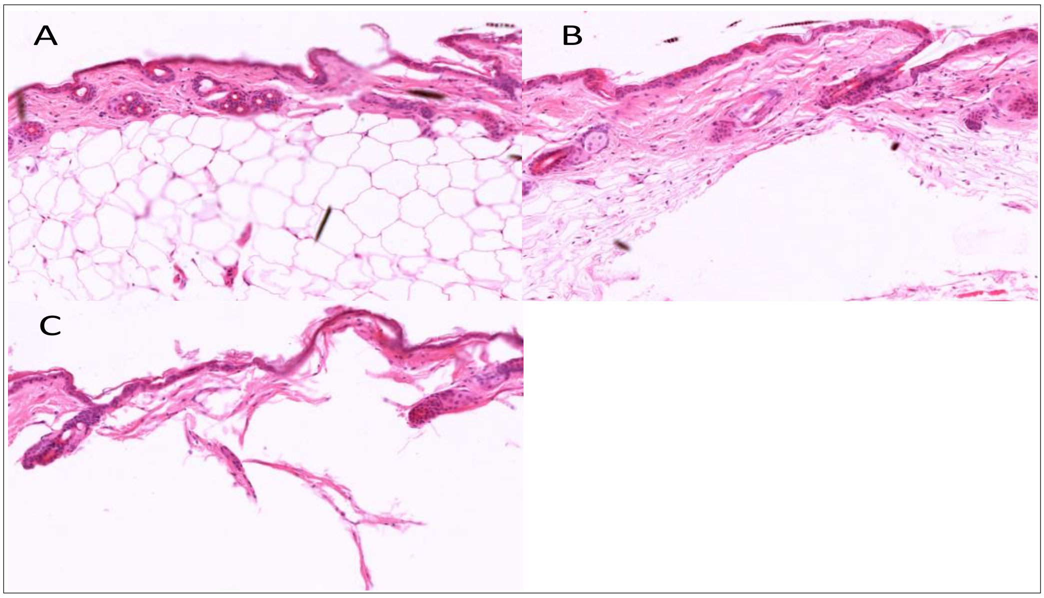


Figure 4.22 Stain error of H&E imagae of mice skin. A) Damage in epidermis layer. B) Damage in fat cell layer. C)Damage in in dermis layer

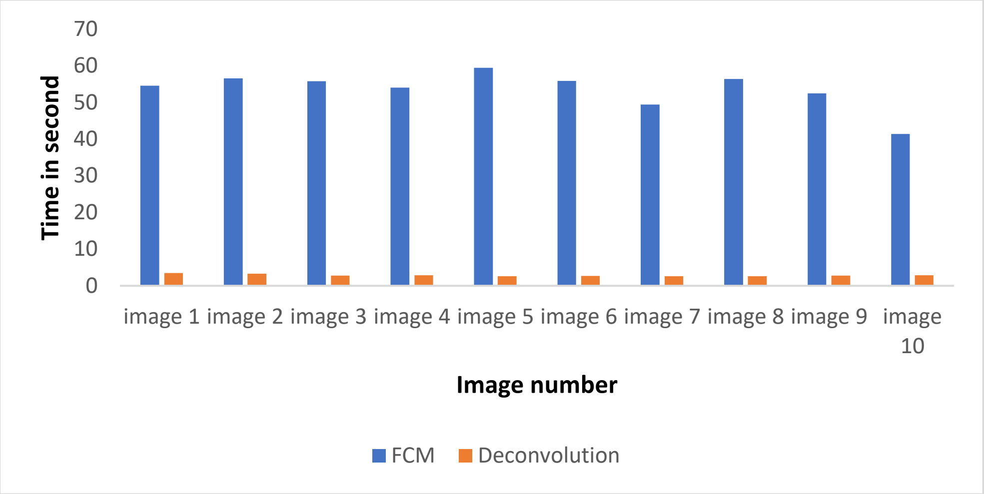


Figure 4.23 Comparison of time consumed between FCM and Deconvolution methods

4.2.3 Adaptive combination of colour deconvolution FCM clustering

As stated earlier, the colour deconvolution seems to work well on simple images whereas the fuzzy C-mean clustering works well on more complex ones. The section proposes an automatic method to check the histogram of the red channel of the image for the number of peaks. If it there is less than two peaks, the image will be automatically fed into the colour deconvolution method. Otherwise, it will be forwarded to the fuzzy C-mean clustering as shown in Figure 4.24.

The number of peaks of the image were determined automatically depending on the image histogram, where the three parameters have been used to determine the peaks are prominence, width and distance, whose values were 40, 2 and 20 respectively. The prominence (length of peak) is selected to have a vertical drop of more than 40 from the peak on both sides without encountering either the end of the signal or a larger intervening peak. The width is selected at 2, in order to find peaks whose width is at least 2. Finally, the distance between peaks is set at 20, in order to ignore smaller peaks that may occur in close proximity to a large local peak.

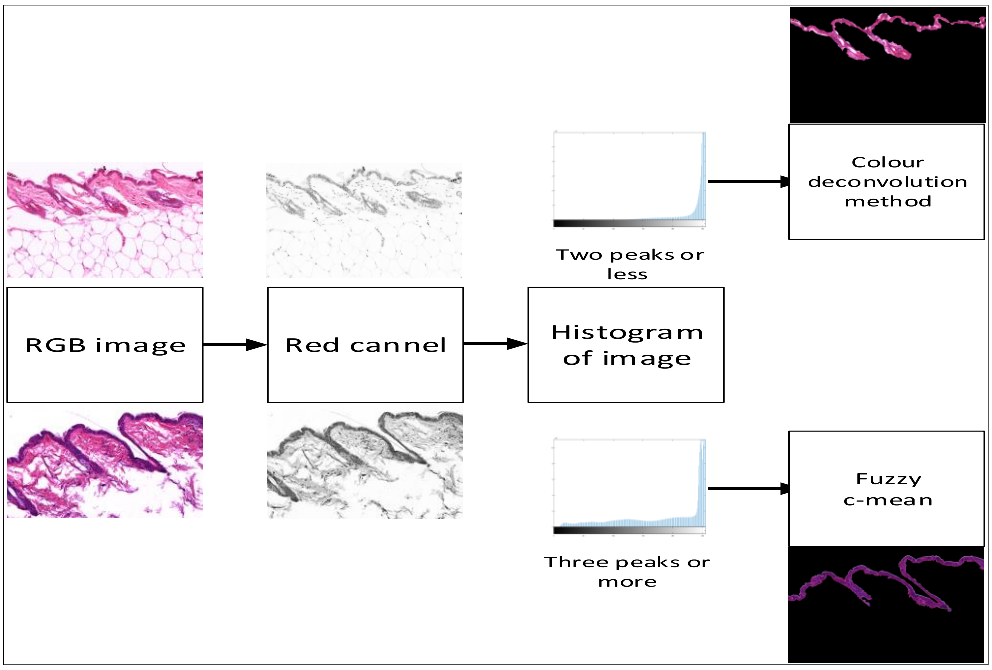


Figure 4.24 Automatic fusion segmentation method based on the histogram of image, where fuzzy represents the method for the image that has three or more peaks and the colour deconvolution method for the image which has two or less peaks

4.3 Experiments and Results

H&E stained skin images from 16-week-old female genetically matched wild-type (30) animals and individual genotype (462 different gene knock-out) animals were used. The development data set contains 7,000 H&E stained images at 20X magnification and a resolution of 1444X908 in a Tagged Image File Format (TIFF). The design of the MGP pipelines (use of diets and genetic backgrounds etc. is beyond the scope of this project but can be reviewed in (Collins *et al.*, 1998). Our results consist of three parts, the first part shows the colour deconvolution method results. Second part display the results of FCM technique and the third part present the results of fusion contrubtion method. We used MATLAB to impement all our contribtuions in all chapters and we used desktop computer dell i5, with 1 T hard derive.

4.3.1 Results of Colour Deconvolution Method

The methods explained in this chapter were applied on the 7,000 H&E images described above. However, as the manual verification of all images by a domain expert is a time-consuming process, we decided to evaluate the results by a random sampling in 7 rounds, where in each round 100 images were randomly selected from each 1,000 images. The boundaries of binary masks of the output were superimposed on the original images and then manually establish the correctness of the segmentation by a domain expert. The final segmentation output for three main skin layers and the epidermis layers from a representative H&E images is illustrated in Figure 4.25 and the accurate of the layers segmentation was shown in the tables (4.1- 4.4).

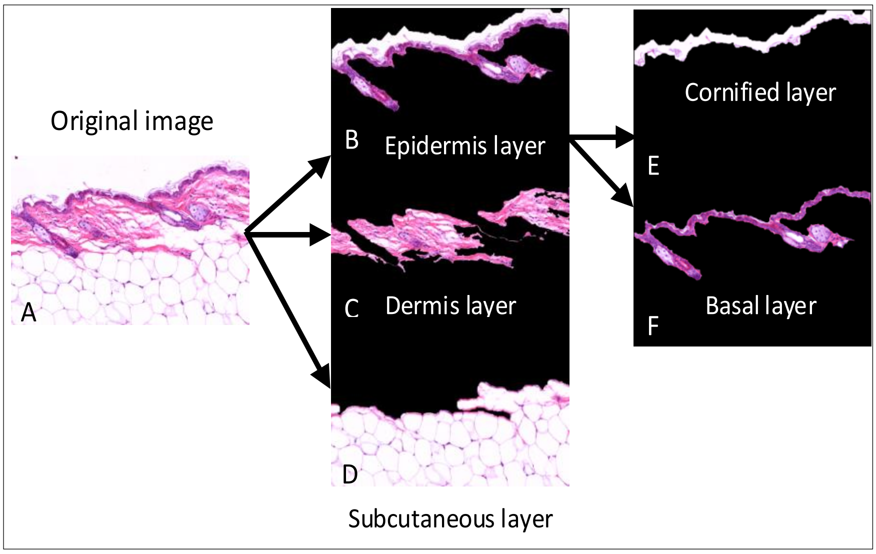


Figure 4.25 Layer segmentation of H&E images. A) Original image. B) Epidermis layer segmentation. C) Dermis layer segmentation. D) Adipose layer segmentation

Table 4.1 Results of a seven round sampling experiment for segmentation of the basal layer

Epidermis layer (basal layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	75	20	5
Round2	66	25	9
Round3	62	28	10
Round4	69	23	8
Round5	81	14	5
Round6	65	24	11
Round7	74	18	8
Average	**70.28**	**21.71**	**8**

Table 4.2 Results of a seven-round sampling experiment for segmentation of the cornified layer

Epidermis layer (cornified layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	71	25	4
Round2	68	24	8
Round3	66	22	12
Round4	64	30	6
Round5	73	20	7
Round6	77	17	6
Round7	80	15	5
Average	**71.28**	**21.85**	**6.85**

Table 4.3 Results of a seven round sampling experiment for segmentation of the adipose layer

Adipose layer			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	69	20	11
Round2	73	18	9
Round3	67	25	8
Round4	75	16	9
Round5	77	17	6
Round6	63	32	5
Round7	78	14	8
Average	**71.71**	**20.28**	**8**

Table 4.4 Results of a seven-round sampling experiment for segmentation of the dermis layer

Dermis layer			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	73	19	8
Round2	82	11	7
Round3	60	30	10
Round4	66	20	14
Round5	69	21	10
Round6	62	29	9
Round7	75	18	7
Average	**69.57**	**21.14**	**9.28**

Each experiment selected 100 images randomly from 7,000 H&E images. The accurate segmentation in the tables (4.1- 4.4) indicate perfect split of layers, particularly (70.28%) for accuarte segmentation in the epidermis layer (basal layer) as shown in Figure 4.25F and in the table (4.1), (71.28%) for accuarte segmentation in epidermis layer (cornified layer) as shown in Figure 4.25E and in the table (4.2), (71.71%) in the adipose layer (subcutaneous laye) as shown in Figure 4.25D and in the table (4.3), (69.57%) in the dermis layer as shown in Figure 4.25C and in the table (4.4).

While the part of mask is not fully accurate, meaning that the layer were segmented only partially as shown in Figure 4.27B (i.e. not a complete mask, just a part of the layer), as shwon in table (4.1) where the value (21.71%) indicates to part of the layer in the epidermis layer (basal layer), the value (21.85%) in the table (4.2) indicates to part of the layer in the epidermis layer (cornified layer), the vaule (20.28%) as shown in table (4.3) indicates to part of the layer in the adipose layer and the value (21.14%) as shown in the table (4.4) indicates to part of the layer in the dermis layer.

There was also an incorrect segmentation of layers, which is wrong layer segmentation as shown in Figure 4.27A and in the table (4.1) where the value (8%) indicates to wrong layer segmentation in the epidermis layer (basal layer), where the value (6.85%) indicates to wrong layer segmentation in the epidermis layer (cornified layer), where the value (8%) indicates to wrong layer segmentation in the adipose layer, where the value (9.28%) indicates to wrong layer segmentation in the dermis layer. We have compared the masks of the results manually with the original images to determine an accurate result. As a result, the incorrect segmentation was determined, the images and H&E staining of the skin was found to have issues. For example, there were some slides which had issues in relation to illumination and non-uniform illumination problems can affect RGB values and affect their comparability with other H&E in the dataset (Zhang *et al.*, 2011). One other example of errors is from damage to the tissues, which can yields inaccurate segmentation.

Figure 4.26 illustrates a damaged dermis when compared to a normal skin image in B

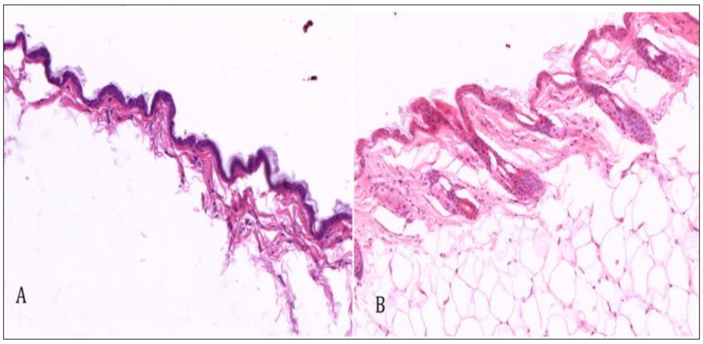


Figure 4.26 H&E images. A) Damage snapshot. B) Non-Uniform illumination image

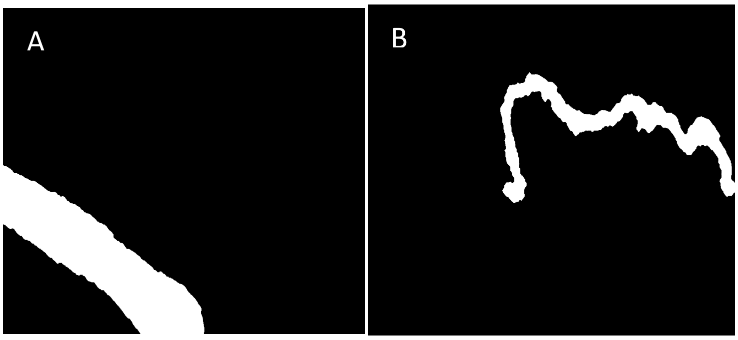


Figure 4.27 Segmentation errors. A) Incorrect segmentation. B) Part of mask

4.3.2 Results of FCM Method

We used the image feature and image number as mentioned before. The result was achieved by implementing randomly 7 rounds, where in each round 100 images were randomly selected from 1000 images as shown in the following tables (5-8).

Table 4.5 Results of a seven-round sampling experiment for segmentation of the basal layer by using fuzzy method

Epidermis layer (basal layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	85	10	5
Round2	95	3	2
Round3	93	4	3
Round4	96	2	2
Round5	90	6	4
Round6	94	2	2
Round7	97	2	1
Average	**93**	**4**	**3**

Table 4.6 Results of a seven-round sampling experiment for segmentation of the cornified layer by using fuzzy method

Epidermis layer (cornified layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	89	5	6
Round2	95	2	3
Round3	88	5	7
Round4	82	10	8
Round5	96	3	1
Round6	95	4	1
Round7	90	4	6
Average	**90.71**	**4.71**	**4.57**

Table 4.7 Results of a seven-round sampling experiment for segmentation of the dermis layer by using fuzzy method

Dermis layer			
Experiences	**Accurate segmentation**	**Part of mask (%)**	**Incorrect segmentati**
Round1	93	4	3
Round2	88	8	4
Round3	90	6	4
Round4	92	3	5
Round5	87	10	3
Round6	91	6	3
Round7	95	3	2
Average	**90.85**	**5.71**	**3.42**

Table 4.8 Results of a seven-round sampling experiment for segmentation of the adipose layer by using fuzzy method

Adipose (fat cell) layer			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	96	3	1
Round2	93	5	2
Round3	89	9	2
Round4	95	3	2
Round5	90	7	3
Round6	97	2	1
Round7	95	3	2
Average	**93.57**	**4.57**	**1.85**

As in the experiments in section 5, the 100 images were selected randomly from 7,000 H&E images of our data set, as shown in tables 4.5, 4.6, 4.7 and 4.8. The accurate segmentation meant that the layers of skin had parted perfectly, while the correct segmentation meant that the layer had split, but it did not include a complete mask, just part of the layer, due to some pixels beyond to the wrong cluster. Where the incorrect segmentation indicates segmentation of the wrong layer. The comparison was then achieved manually between the original H&E images and the automatic layer segmentation mask to assist the accurate result.

4.3.3 Results of Adaptive Combined Method

The same 100 images used above were used again in this section. The hybrid method is fusion method between the colour deconvolution and FCM method as shown in the following tables (4.9-4.12).

Table 4.9 Results of a seven-round sampling experiment for segmentation of the epidermis layer (basal layer) by using fusion method

Epidermis layer (basal layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	95	3	2
Round2	97	2	1
Round3	97	3	1
Round4	96	3	1
Round5	95	3	2
Round6	98	1	1
Round7	97	2	1
Average	**96.42**	**2.42**	**1.28**

Table 4.10 Results of a seven-round sampling experiment for segmentation of the epidermis layer (cornified layer) by using fusion method

Epidermis layer (cornified layer)			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	98	1	1
Round2	97	2	1
Round3	98	1	1
Round4	98	1	1
Round5	96	3	1
Round6	97	3	1
Round7	97	3	1
Average	**97.282**	**2**	**1**

Table 4.11 Results of a seven-round sampling experiment for segmentation of the dermis layer by using fusion method

Dermis layer			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	95	4	1
Round2	96	3	1
Round3	94	4	2
Round4	97	2	1
Round5	96	4	2
Round6	98	1	1
Round7	95	3	2
Average	**95.85**	**3**	**1.42**

Table 4.12 Results of a seven-round sampling experiment for segmentation of the adipose layer by using fusion method

Adipose (fat cell) layer			
Experiences	**Accurate segmentation (%)**	**Part of mask (%)**	**Incorrect segmentation (%)**
Round1	98	1	1
Round2	96	2	2
Round3	97	2	1
Round4	96	3	1
Round5	95	4	1
Round6	97	2	1
Round7	97	2	1
Average	**96.57**	**2.28**	**1.14**

As shown the results of the adaptive combination method (%96.42 for basal layer, %97.282 for cornified layer, %95.85 for dermis layer, and %96.57 for adipose) are significantly better than the results of using FCM and colour deconvolution method, because it address the FCM shortcoming issues. Although the FCM method is better than the colour deconvolution method but in some cases such as (10 out of 100 images) the FCM method segment a part of layer while the colour deconvolution segment a complete layer as shown in Figure 4.28. Because of the segmentation in FCM sometimes causes loss of a part of mask layer such epidermis layer, due to some pixels beyond to the wrong cluster as shown in Figure 4.28C, where the image D show correct segmentation by colour deconvolution method. Despite three peaks in image B in Figure 4.28, the colour deconvolution do correct segmentation as shown in Figure 4.28D

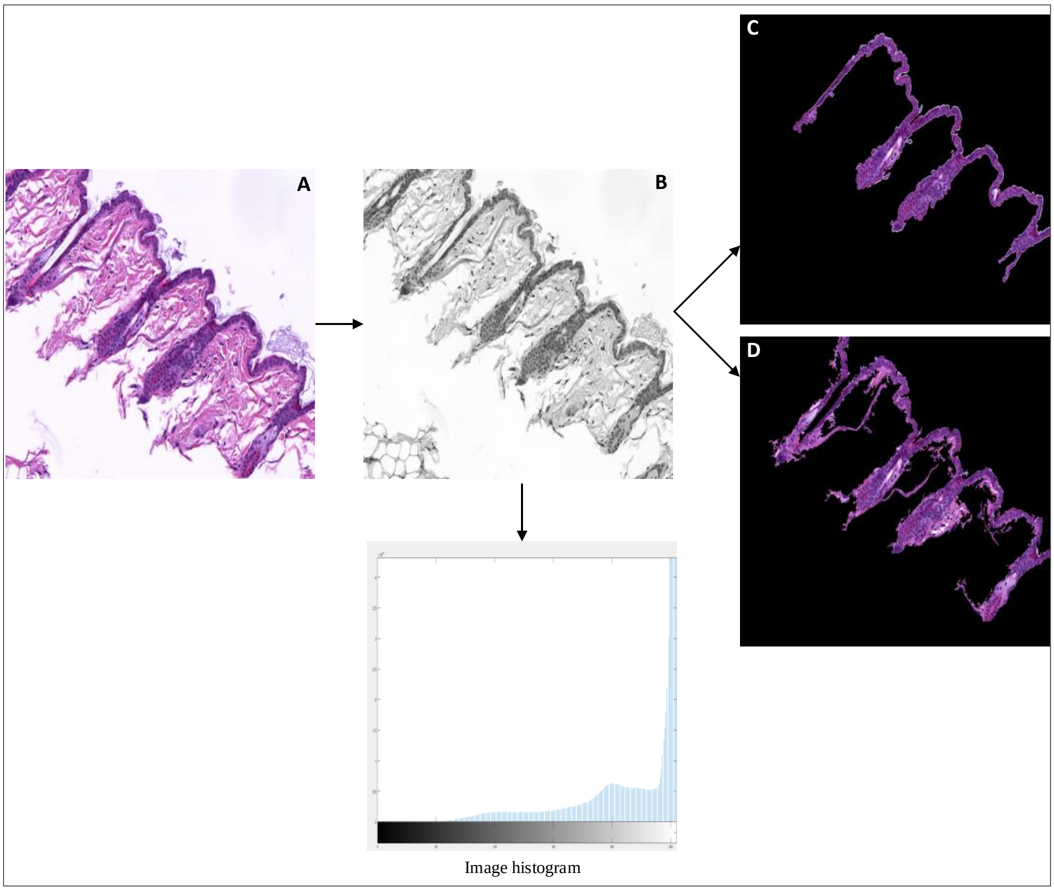


Figure 4.28 Fusion segmentation method. A) Colour image. B) Red channel of image A. C) Epidermis layer segmentation using FCM. D) Epidermis layer segmentation using colour deconvolution method

4.4 Conclusion

In this chapter, we designed and developed an adaptive system that uses one of two automatic methods, (the colour deconvolution and fuzzy c-mean methods) depending on parameter that measures the texture complexity of the images in terms of the number of peaks in the histogram of the red channel. The automatic solutions aimed to segment and sub-segment the three main layers of mice skin with H&E images, namely the epidermis, dermis and fat cell layers (Figure 4.29). We showed that this adaptive hybrid solution that combines the efficiency of colour deconvolution with the robustness of fuzzy C-mean clustering works well on images with different texture complexity and outperforms the deconvolution and fuzzy c-mean methods individually. The chapter also demonstrated the use of different kinds of thresholding such as the Otsu and the inter-modes to overcome the challenges of colour overlapping between the epidermis and the layers are far from satisfactory. The following benefit from the success of segmentation proposed in this chapter, by enabling the quantification of different features/objects associated with individual skin layers/sublayers as a mean of finding underlying genetic basis of some diseases.

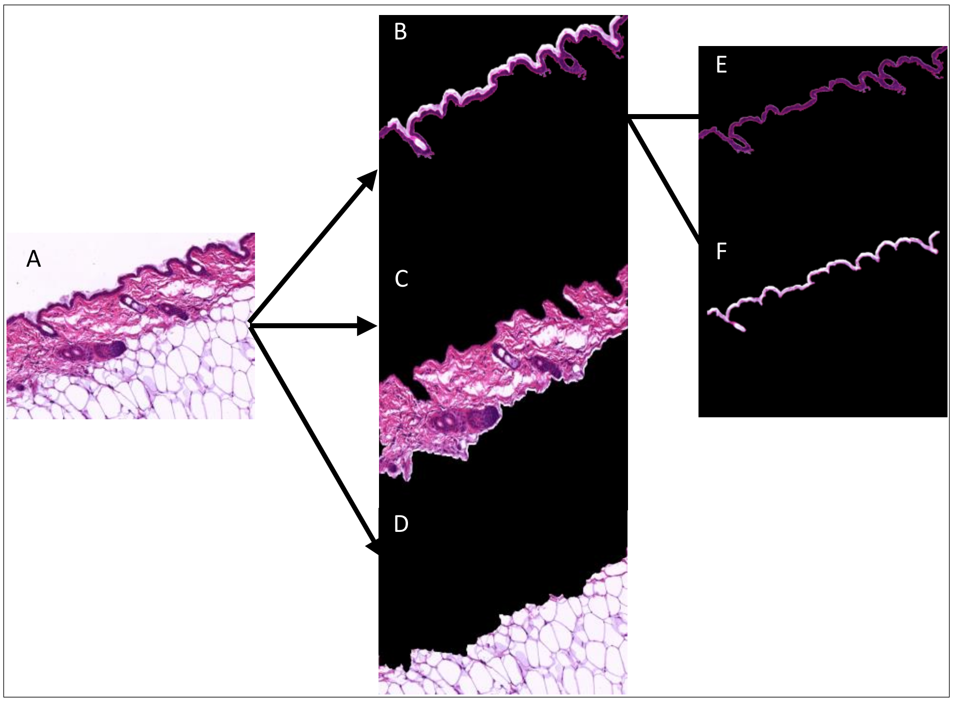


Figure 4.29 Segmentation of three main layer in the mice skin in H&E images by using fuzzy c-mean method. A) Colour H&E image which consist of three main layers. B) Epidermis layer segmentation. C) Dermis layer segmentation. D) Fat cell layer segmentation. E) Basal layer sub-segmentation. F) Cornified layer sub-segmentation

Chapter 5 : Automatic Nuclei Segmentation Based on Fuzzy C-Mean

Quantifying changes to the shape and number of nuclei in skin tissue is essential for assessing the state of some genetic-based disease situation, such as skin cancer. Observed changes in the distribution of nuclei clusters and the shape of the nuclei in different layers can be caused by certain skin diseases that can actively or passively control their shape and/or positions (Brandt *et al.*, 2006; Melcer and Gruenbaum, 2006; Huber and Gerace, 2007; Edens *et al.*, 2013). Moreover, the gene knockout of mice skin image has made it possible to detect the genetic basis of some phenotypes in different layers. This chapter exploit the results of skin layers segmentation to facilitate the identification and counting the number of nuclei in two layers: the epidermis layer and dermis layer.

Traditional manual methods to nuclei counting is time-consuming and could have notable inter-observer variabilities due to the subjective decisions made by different manual annotators. Therefore, there is a need to automate systems for nuclei counting in different skin layers that are reproducible results. This chapter has three main contributions:

1. Develop an automated nuclei segmentation method in the epidermis layer that overcomes the overlapping of nuclei problem.
2. Propose an automatic nuclei segmentation scheme in the dermis layer.
3. Use the nuclei counting schemes for a dataset of wild-type and mutant images for identifying genes causing changes in nuclei count, in relation to skin abnormalities.

The rest of the chapter is organised as follows. In Section 5.1 we describe the problem of segmenting nuclei in mouse skin images at different layers. In Section 5.2 algorithm framework of our approach is presented. Section 5.3 evaluate the performance of our automatic algorithms and in comparison with manual segmentation and present experimental results on identifying genes causing skin abnormalities from nuclei counts. Conclusions of the contributions were described in Section 5.4.

5.1 Introduction and Problem Statement

The recent development of high-throughput production of targeted, genetically altered mouse lines has enabled a systematic phenotypic screening and gene discovery.

Identifying genes correlated with changes in nuclei in different layer is a key step towards understanding the genetic basis of skin cancer.

Accurate identification of any changes in nuclei is of a significant biological and clinical relevance. For example, mutations in the PTCH1 and SMO genes in human have been associated with basal cell carcinoma which is a type of skin cancer that develops in the epidermis layer (Marzuka and Book, 2015) (Lang et al., 2019; Verkouteren et al., 2019). Figure 5.1, below, illustrates abnormal types of skin tumour linked with groups of nuclei throughout the epidermis while showing a normal epidermis.

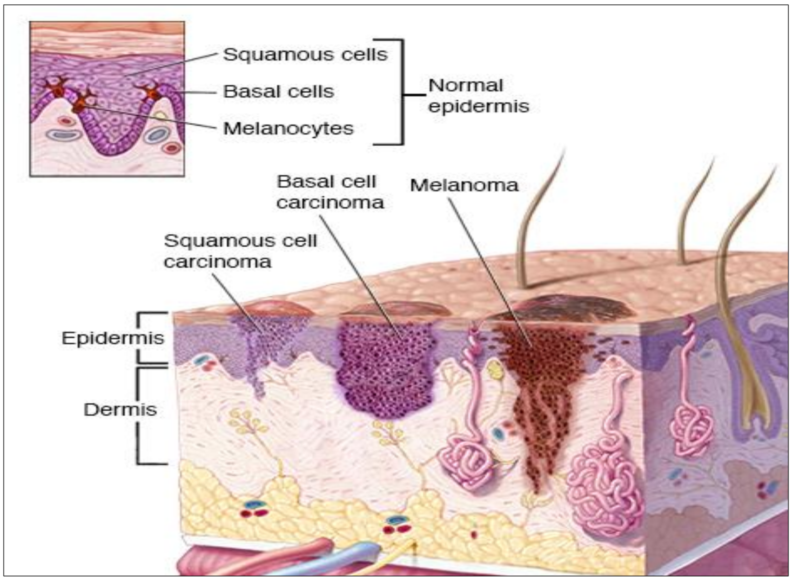


Figure 5.1 Types skin cancer. Melanoma and basal cell carcinoma (Findlay and Ally, 2015)

Melanoma has also been reported as another type of skin cancer that develops in the dermis layer with mutations in the BRAF and NRAS genes (McConnell *et al.*, 2019). That causes a tumour in the dermis, lead to increasing the nuclei in this area, also illustrated in Figure 5.1. It is likely that there are further, unidentified genes in which mutations may also cause a skin cancer. Therefore, the identification of genes associated with nuclei number in different layers may provide insight into both specific and generalised skin abnormalities**.**

H&E stain technologies that used to show the skin details such as nuclei, using chemical staining method in histopathology. Due to huge variability in various chemical formulations, thickness, lab protocols and different image scanners, automatic image analysis method of microscopy is prone to errors (Bayramoglu *et al.*, 2017). In addition, accurate measurement of nuclei from typical H&E stain, is adversely effected as a result of overlapping of colour intensity between nuclei and surround area as illustrated in Figure 5.2.

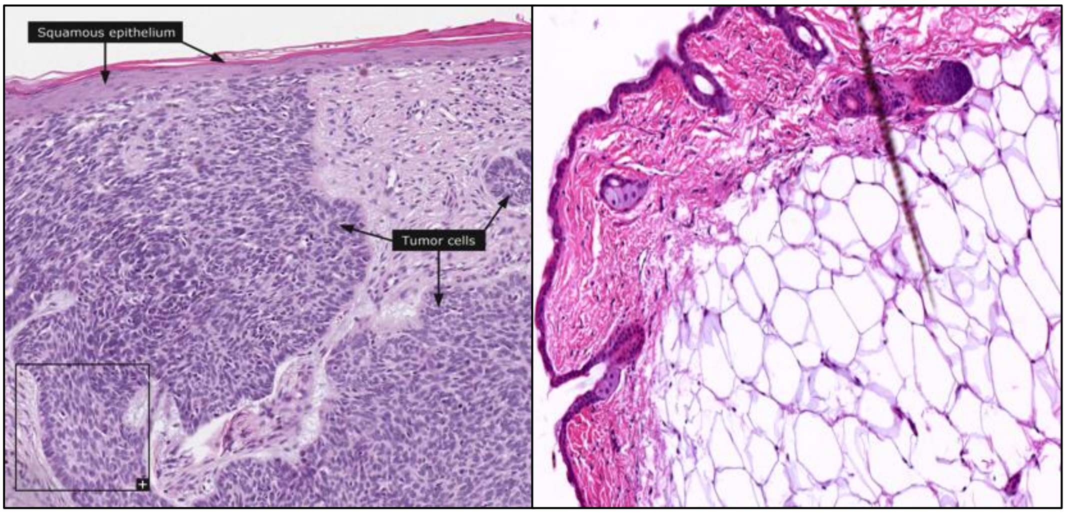


Figure 5.2 Microscopic H&E image. Microscopic human image (Mathias et al., 2015) (on the left). Microscopic mice image (on the right)

While it is clear that abnormality in the human image, as shown in Figure 5.2 (on the left side), is clearer than in the mice images. The tumour area in the human tissue image is highlighted by the group of nuclei with almost constant intensity and similar shape, making the task of segmenting the tumour that much easier. Whilst in the case of mouse skin image the nuclei have almost the same intensity with the background area or surround texture, and therefore human annotation while examining the images under a microscope is used to determine changes in the number of nuclei and to diagnose tumour abnormalities.

In general, automatic segmentation of nuclei in different mouse skin layers is a technically challenging task due to difficulties in locating related and shape within the image caused by colour intensity overlapping of different tissues such as blood cells and vessels, etc. In particular, automatic nuclei segmentation in the epidermis layer is difficult due to the potential overlapping of nuclei colour intensity with the surrounding tissue. Another challenge is that of overlapping between nuclei, which lead to more than one nuclei being identifies as the same object, which require schemes to separate them. All these problems are expected to have impact on the accuracy of nuclei counting.

The challenge of segmenting the nuclei in the dermis layer is somewhat different in that separable objects in the dermis could include nuclei, and other objects such as blood cells/vessels, etc.

However, nuclei can be categorized from the rest of these objects in terms of their near circular shape. These added challenges must be addressed, besides all pervious issues, when designing fully automatic solutions.

All the aforementioned challenges do not only impose difficulties for machine-based solutions but even for expert analysts with whom inter- and intra-analyst variations in manual measurement results could happen. In turn, this leads to difficulties in obtaining a solid ground truth on manual measurements to objectively assist the evaluation of the automatic solutions. Given the volume of images that are generated by the high-throughput screening, an automated solution for annotation and triaging of images for human analysis will not only reduce costs in time and human resources, but also eliminates the likelihood of human errors and inconsistency in manual measurements.

5.2 Algorithmic framework of Nuclei Segmentation

Our proposed method of counting nuclei in the dermis and epidermis layers will adopt the use fuzzy cluster techniques (Xiao and Peng, 2013; Bizrah et al., 2014; Bibiloni, González-Hidalgo and Massanet, 2016; Radojević, Smal and Meijering, 2016; Dias et al., 2017). Accordingly, after layers segmentation method is applied to segment the epidermis and dermis layers, as explained in Chapter 4, our method segment the nuclei in the epidermis and dermis layers is completed by six key procedures in Pseudocode 5.1:

```
1- Segmenting the nuclei in the epidermis layer using fuzzy cluster method.
2- Identifying the overlapping nuclei in the epidermis layer using circularity
   test and aspect ratio features.
3. Separate overlapping nuclei in the epidermis layer by using watershed with
   distance transform method.
4- Segment the nuclei in dermis layer using fuzzy cluster method.
5- Refine the initial segmentation using shape feature extraction.
6- Count the nuclei in two layers epidermis and dermis.
```

Pseudocode 5.1 Steps for nuclei segmentation of dermis and epidermis layer

Figure 5.3 illustrates the main stages of the proposed framework.

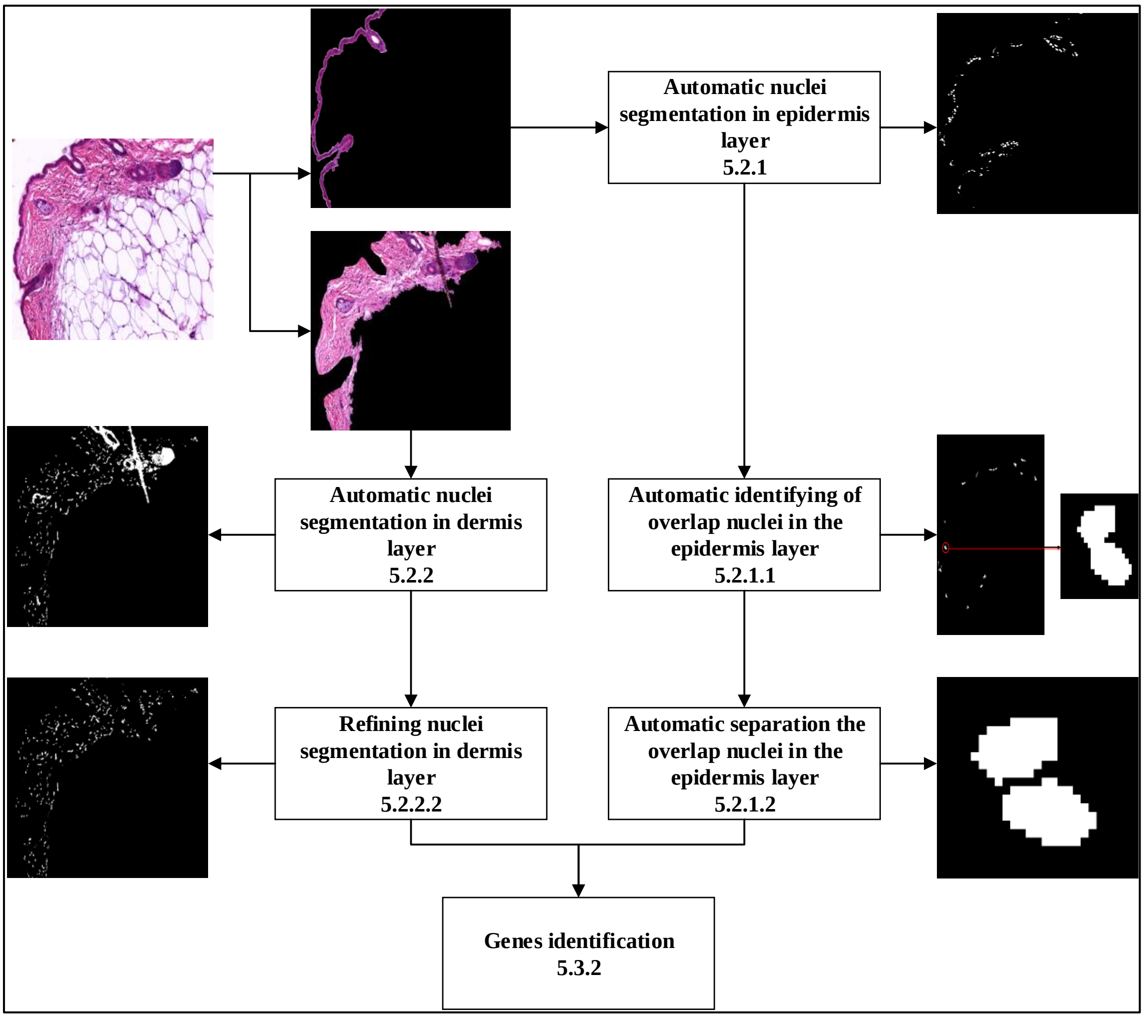


Figure 5.3 Main procedures of proposed method

The following sections describe details of the processes in the above 6 stages.

5.2.1 Nuclei Segmentation in the Epidermis Layer

Segmenting the nuclei in the already segmented epidermis layer works in four stages:

Stage 1. Start by applying fuzzy c-mean clustering on the image red channel, to determine the best cluster (see Chapter 4, Section 3.2). See Figure 5.4 for illustration clusters. Next apply fuzzy threshold on the best cluster to have two images as shown in Figure 5.5.

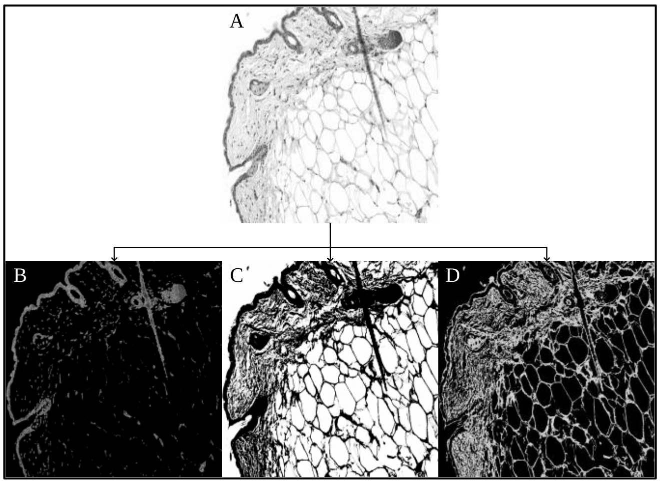


Figure 5.4 Selection the best cluster of output of fuzzy c-mean cluster .A) Red channel of colour image. B) Cluster one (best cluster in our work) which is highlight the epidermis layer. C) Cluster two which is highlight the image background. D) Cluster three which is highlight the texture of all layers (see Chapter 4, Section. 3.2))

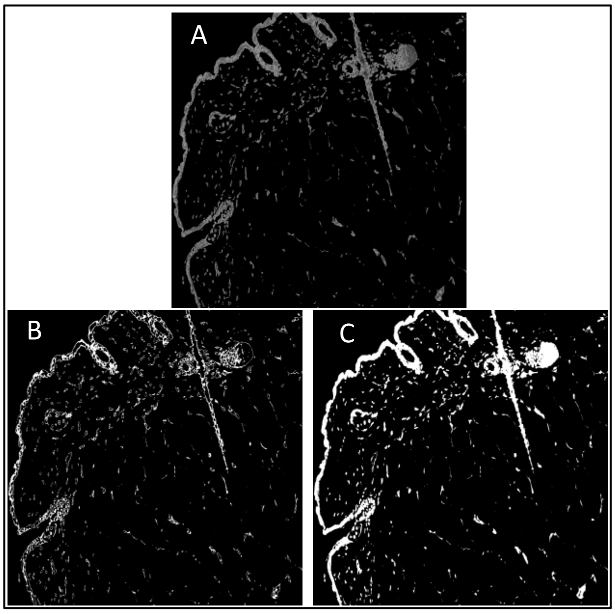


Figure 5.5 Fuzzy c-mean output. A) Particular select cluster from Red Channel. B) Output of FCM threshold (level was selected depend on small and middle data point class (see Chapter 4, Section. 3.2)), which represent the texture of epidermis layer and other nuclei in other layers. C) Output of FCM threshold (selected level depends on middle and large data point class)

Stage 2. Pixel-by-pixel subtract image in Figure 5.5B, which represents the texture of the epidermis layer and other objects, and image in Figure 5.5C which represent the epidermis layer as a solid mask and nuclei in the dermis layer, where the output is shown in Figure 5.6C. The reason for applying the subtraction operation relates to the nature of image in Figure 5.6B has a solid object (epidermis layer mask) while image in Figure 5.6A has the texture or the boundary of the objects (epidermis layer mask). The subtraction operation avoids the border of nuclei and extracting the core object from the solid mask (St-Charles and Bilodeau, 2014; Prajapati and Jadhav, 2015).

Stage 3. Pixel-by pixel multiply image in Figure 5.6D, which shows the epidermis layer only as a solid mask, with image in Figure 5.6E, which clarify the inversion of the output of step 1 (output of subtraction operation). The multiplication operation results in avoiding any object outside the epidermis layer. Following this, erosion morphology (by 1 pixel, with disk structure) is applied on the nuclei, to avoid overlapping among the nuclei, as displayed in image in Figure 5.6F.

Stage 4. Remove noise-like objects (not nuclei) or the error object stain was addressed by learning the machine by labelling all objects in the epidermis layer, and avoiding the noise object by make them a background. After many attempts to determine the noise objects, we found a threshold area for whole objects in the epidermis layer, i.e. we found that each object has an area of more than 150, error stain object or not nuclei as shown in Figure 5.7.

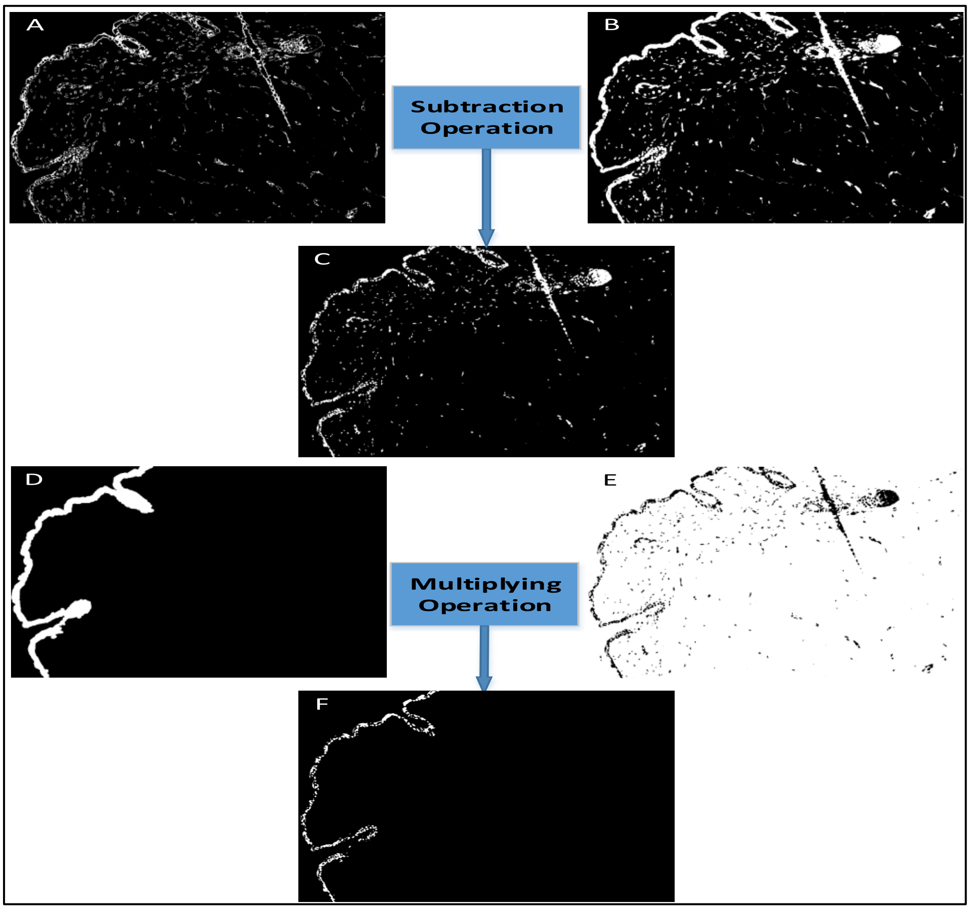


Figure 5.6 Proposal methods for nuclei segmentation. A and B have been explained in Figure 5.5 as B and C respectively). C) The output of the subtraction operation between image A and B. D) Epidermis layer mask using fuzzy c-mean method. E) Inversion of image C, which was the output of the subtraction operation. F) Output of the multiplication operation between image D and E

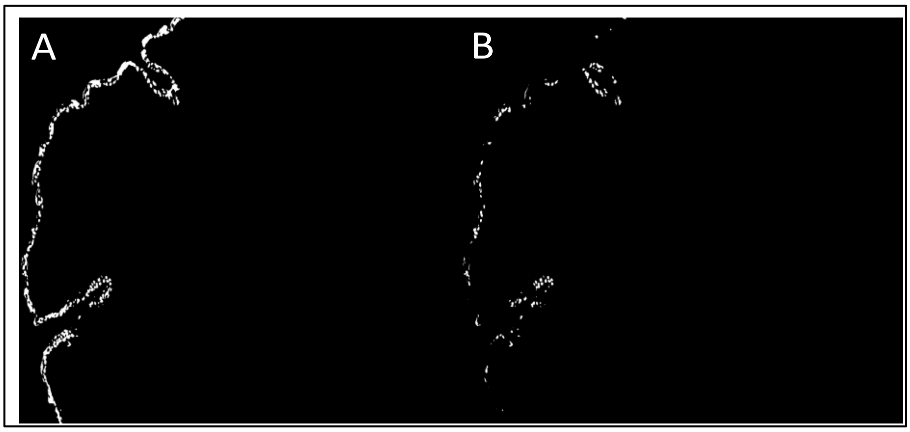


Figure 5.7 Removal of noise objects. A) Epidermis layer before removing the noise objects. B) Epidermis layer after removing the noise objects

5.2.1.1 Identification of overlapping nuclei in the epidermis

Automatic identification of overlapping nuclei of the epidermis layer will be employing the circularity and aspect ratio features. The circularity feature was also employed to determine the relationship between the tumour of lung adenocarcinomas and the feature of computed tomography (CT), which contains the measurements of histogram and texture analysis (Koo *et al.*, 2017). Moreover, in (Guven and Cengizler, 2014), the circularity and aspect ratio parameters were deemed appropriate for classification of overlapping nuclei in Pap smear samples.

The above feature parameters associated with nuclei-like objects are defined by the following equations extracted from an ellipse fitting of the object (see Neal, 2017):

$$\text{Major Axis} = 2*a \qquad \textit{(eq5.1)}$$

$$\text{Minor Axis} = 2*b \qquad \textit{(eq5.2)}$$

$$\text{Aspect Ratio} = a/b \qquad \textit{(eq 5.3)}$$

$$\text{Area} = \pi*a*b \qquad \textit{(eq 5.4)}$$

$$\text{Perimeter} \approx \pi\,(a+b)\left[3\,\frac{(a-b)^2}{(a-b)^2\left[\sqrt{-3\frac{(a-b)^2}{(a+b)^2}+4}+10\right]}+1\right] \qquad \textit{(eq 5.5)}$$

and

$$\text{Circularity} = 4\,\pi*\,\frac{[Area]}{[Perimeter]^2} \qquad \textit{(eq 5.6)}$$

The procedure works on each segmented epidermis image, by determining the mean of circularity and aspect ratio parameters calculated for all the segmented nuclei-like objects in the image, and objects that have circularity and aspect ratio larger than the corresponding means will be declared as an overlapped of more than one nuclei to be separated by the procedure in the next section.

On examining few cases, we found that we actual overlapped nuclei were considered as single nuclei. To avoid this kind of errors, we designed another test to identify the overlapping nuclei. This complementary procedure, repeats the first procedure but replacing the mean of the aspect ratio with the means of two other parameters of the segmented objects: the major axis and perimeter. Again, overlapping was declared when coordinates of array (circularity, major axis, perimeter) were larger than the corresponding means, (see Figure 5.8).

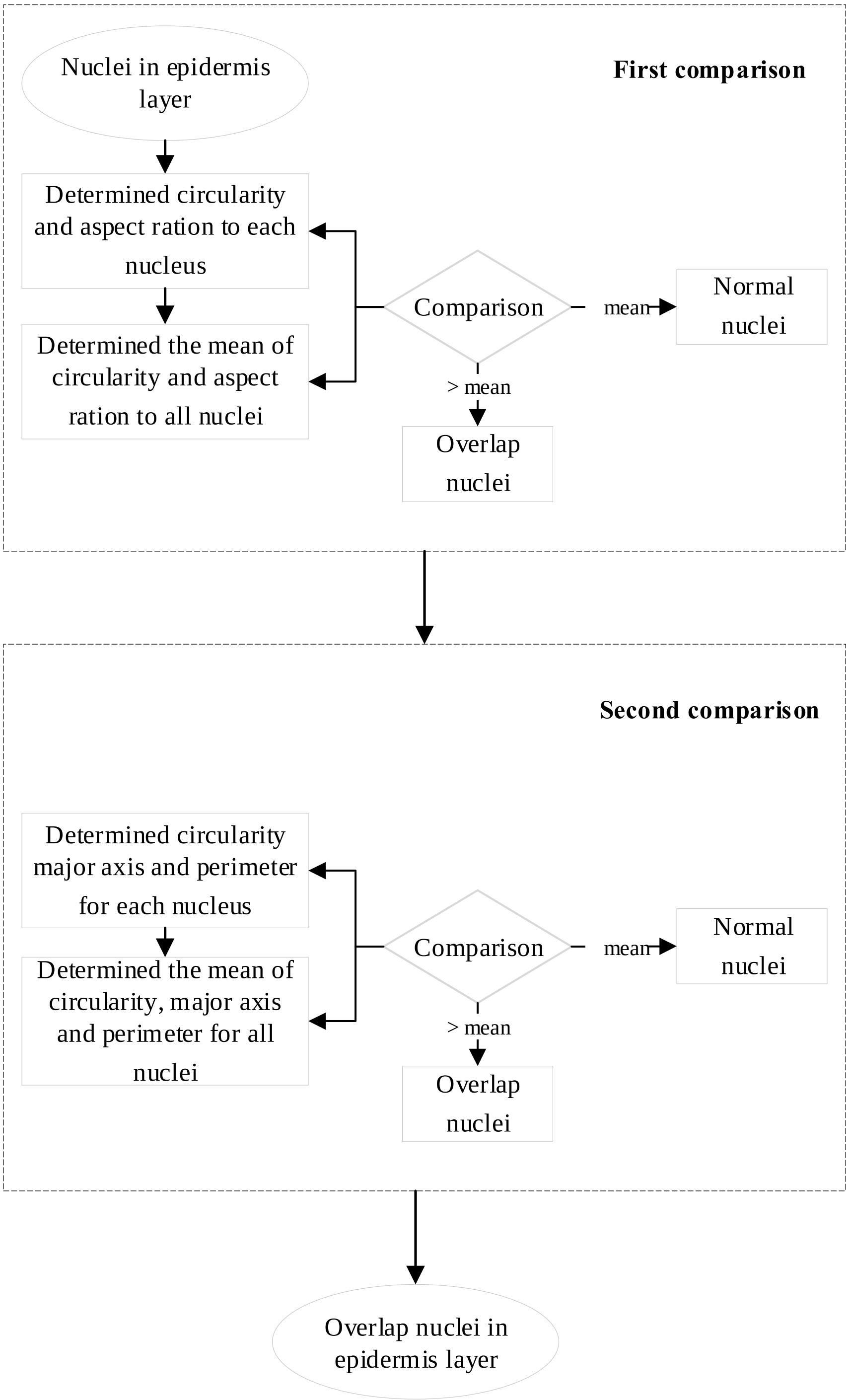


Figure 5.8 Block diagram shows the proposal algorithm of automatically identifying overlapping nuclei in the epidermis layer before performing overlapping nuclei segmentation

Improvement of identifying the overlapping nuclei was achieved by a second comparison, as presented in image of Figure 5.9D, which used the circularity, major axis and perimeter of the overlapping nuclei. This is because those features are sensitive to the length of the object as opposed to the width, where the length of the overlapping nuclei is always longer than the length of single nuclei as displays in Figure 5.10, in the first comparison the aspect ratio was employed, which used the major and minor axes.

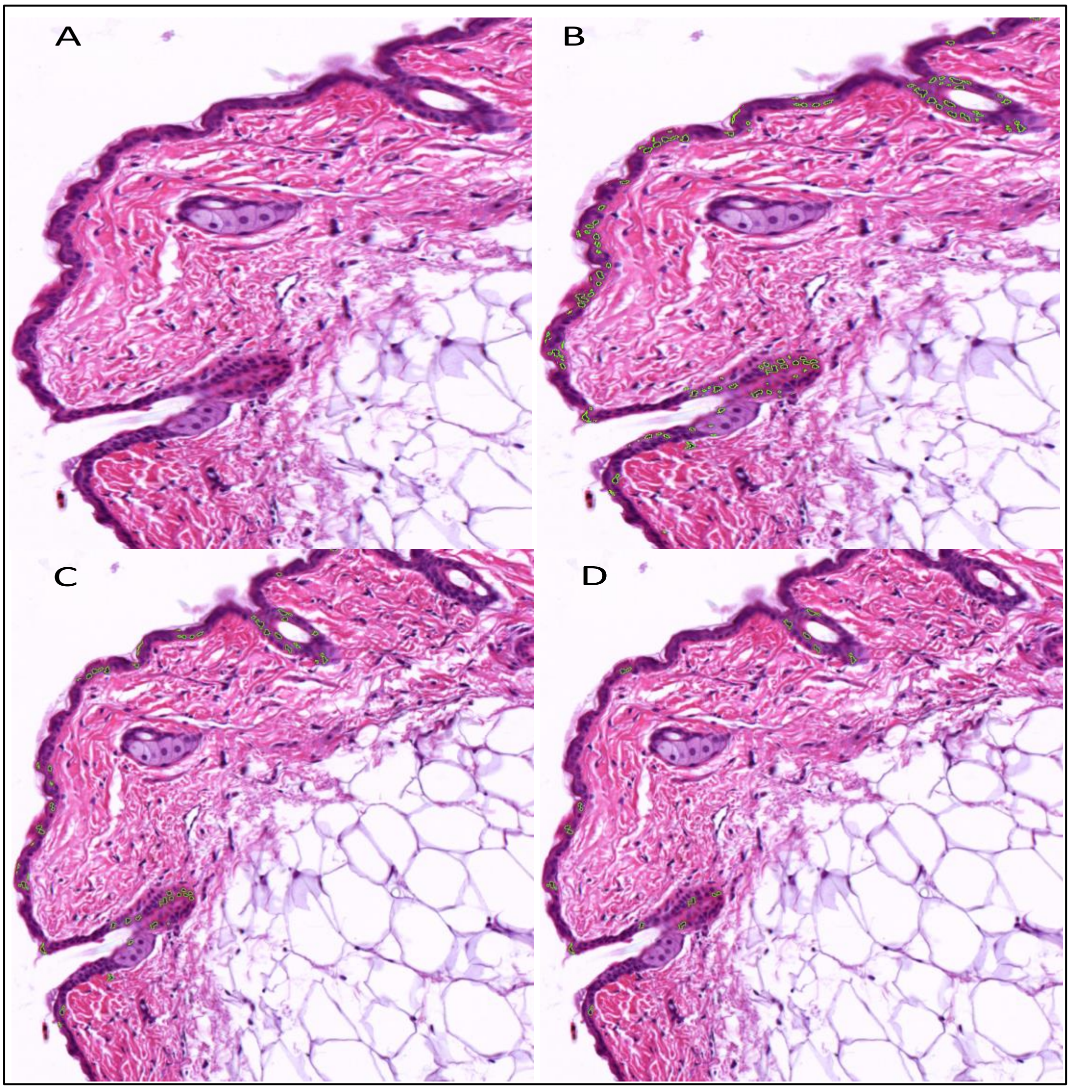


Figure 5.9 Determination of overlapping nuclei. A) H&E original image. B) Nuclei segmentation in the epidermis layer. C) Identification of the overlapping nuclei in the epidermis layer. D) Improving the identification of overlapping nuclei in the epidermis layer

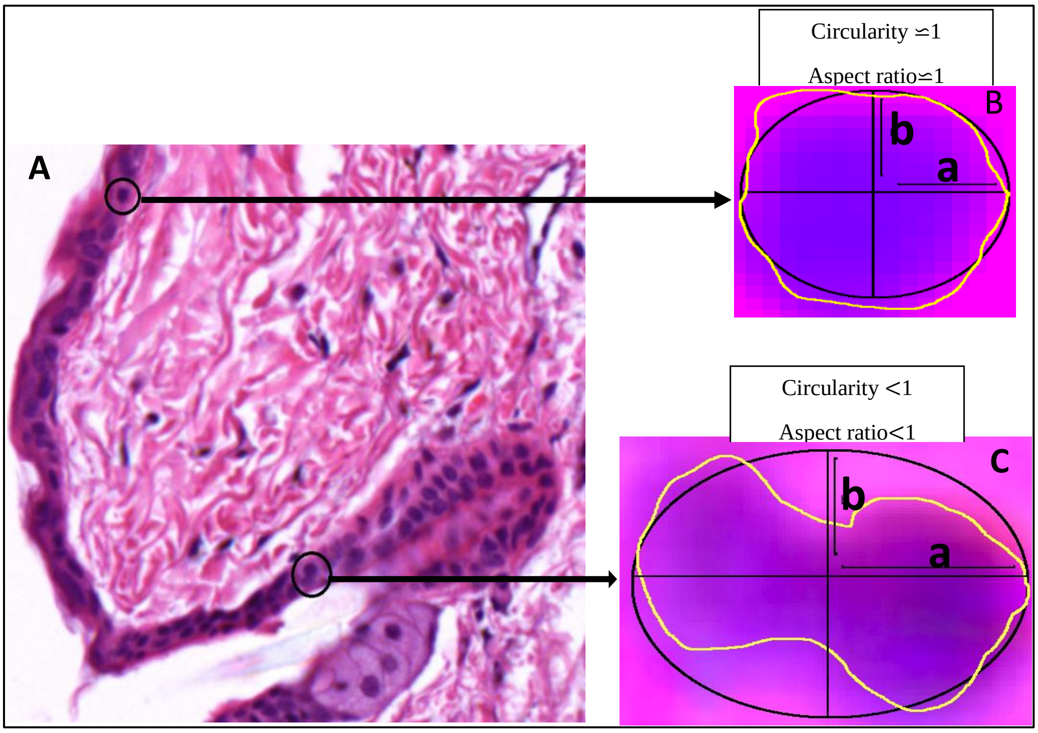


Figure 5.10 Identifying the overlapping nuclei segmentation. A) H&E image with overlapping nuclei and single nuclei. B) Single nuclei segmentation. C) Overlapping nuclei segmentation

5.2.1.2 Separating Overlapped Nuclei in Epidermis Layer

Having identified the overlapping segmented nuclei, as shown in Figure 5.11, it was then required to separate the overlapping nuclei to obtain an accurate counting the nuclei in the epidermis layer. Our search for appropriate overlapping splitting revealed that the watershed associated with the Euclidean distance transform method was the common method. Then H-minima watershed transformation was used for shape marker (Cheng, Rajapakse and others, 2009). The nuclei in the tissue of colon carcinoma was segmented by utilising the region growing technique, then the overlapping nuclei were separated by applying the watershed method based on Euclidean distance (Rogojanu *et al.*, 2010). An automatic method for nuclei segmentation was proposed by using fuzzy combined with the active contour mode, where the fuzzy method provides the initial contour, which led to detection of the contour of nuclei. Touching and overlapping nuclei of breast cancer in the affected tissue were then segmented using the developed watershed method by using a concave vertex, based on Euclidean distance transform (Mouelhi, Sayadi and Fnaiech, 2013). The nuclei of pleural effusion were automatically segmented to diagnose the advanced stages of cancer by using the k-mean cluster method associated with LAB colour space. The overlapping and touching issues was addressed by applying the watershed method based on Euclidean distance transform, where the method of ellipse fitting was used to isolate the boundary of the nuclei

(Win, Choomchuay and Hamamoto, 2017). Accordingly, we deploy this procedure for separating the overlapped nuclei obtained in the previous section. The following is the procedure to segment overlapping nuclei as shown in Pseudocode 5.2:

```
Step 1: Input images with nuclei segmentation in the epidermis layer.
Step 2: Apply distance transform to find the nearest boundary of overlap nuclei.
Step 3: Apply watershed transform to segment the overlap nuclei by set the
        watershed line to zero.
```

Pseudocode 5.2 Steps for overlapping nuclei segmentation

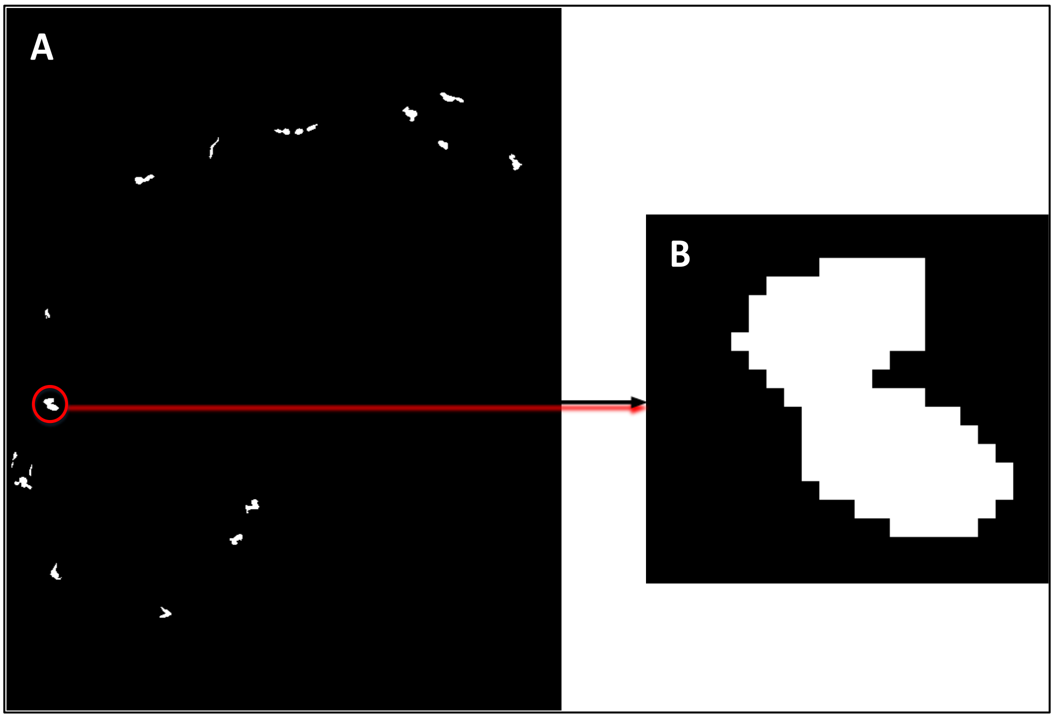


Figure 5.11 Automatic methods for identifying the overlaping method. A) Overlapping nuclei segmentation. B) enlarged sample of overlapping nuclei segmentation

In step 2 in the pseudocode above we using Euclidean distance transform (see Chapter 2) as displayed in Figure (5.12, A, D), the reason for that was to prepare the overlapping nuclei for the next stage which uses the watershed transform by making make two catchment basins in each overlapping nuclei, one basin for each nucleus.

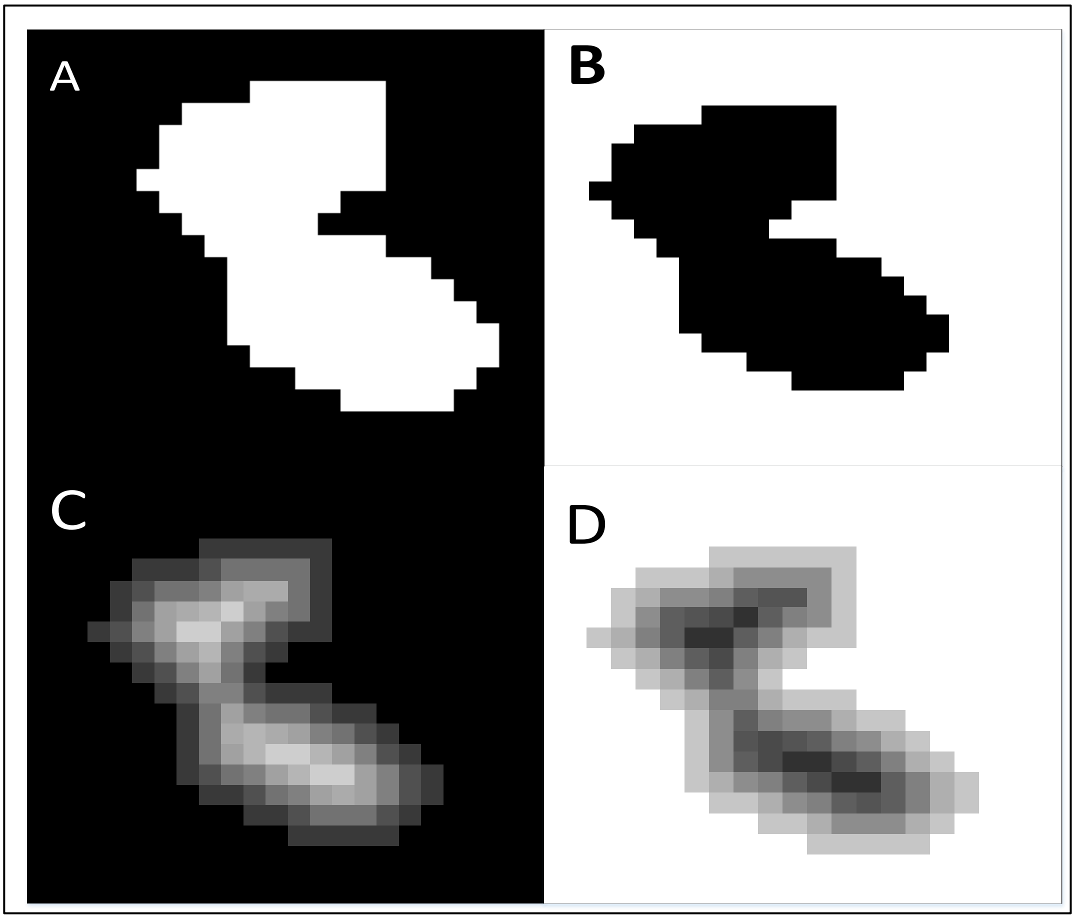


Figure 5.12 Euclidean distance transform of overlapping nuclei. A) Binary overlapping nuclei. B) Inversion of binary overlapping nuclei. C) Calculation of Euclidean distance transform. D) Complement of distance transforms

In step 3 we using the output of distance transform in step2 as input to the watershed transform (see Chapter2) which segment two basins of overlapping nuclei as shown in Figure 5.13, by setting the watershed line (Figure 5.14) to zero, which is located between the two basins.

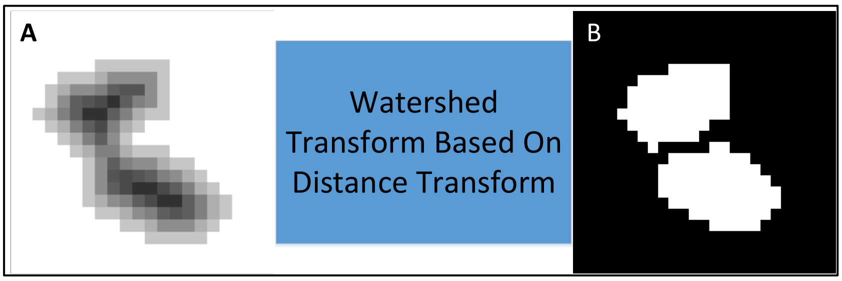


Figure 5.13 Segment the overlapping nuclei. A) Image of distance transforms. B) Segmentation of overlapping nuclei

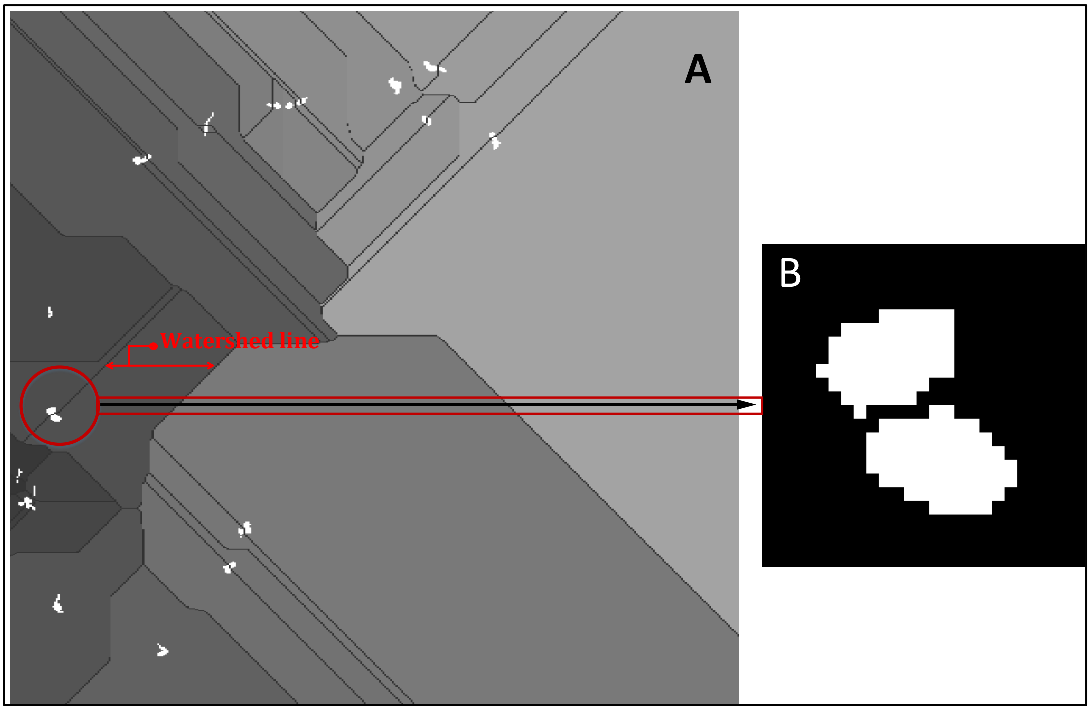


Figure 5.14 Watershed map. A) Two images combined the watershed image and nuclei segmentation. B) Segmentation of overlap nuclei

In some cases two types of errors were found, the first error was that the overlapping were nuclei still not separated because of the watershed and distance transform doesn't recognise the object as two nuclei, such as the nuclei in Figure 5.15A, where the algorithm doesn't segment the overlapping nuclei which is quite evidently two nuclei, based on ground truth as viewed in Figure 5.15C. The second error was the wrong segmentation, for example in Figure 5.15B the overlapping object has four nuclei based on ground truth, and after segmentation the overlapping will be five nuclei as shown in Figure 5.15A) which is incorrect.

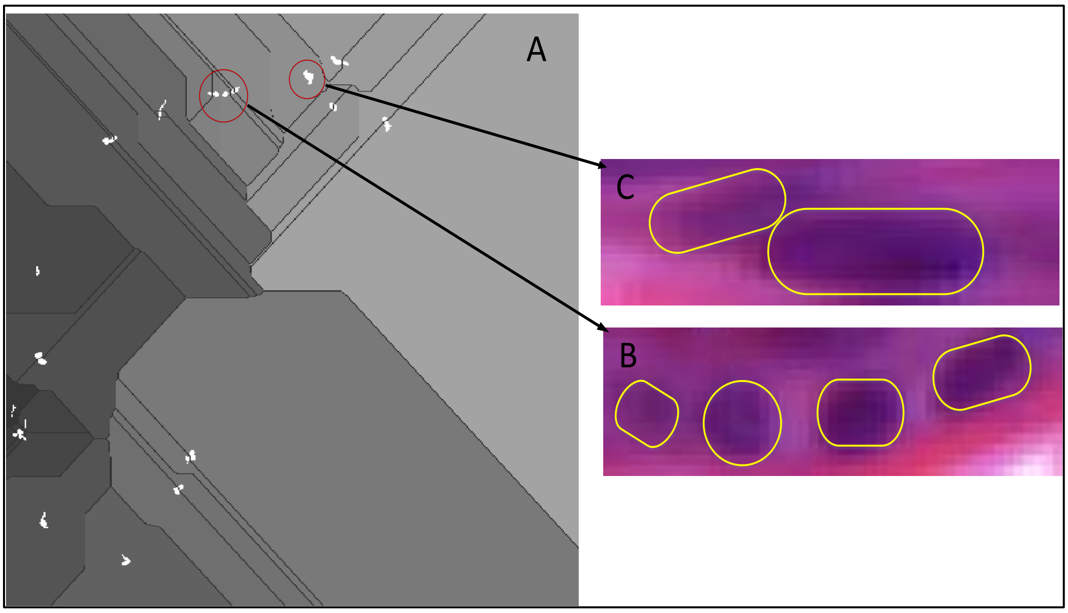


Figure 5.15 Error segmentation of overlapping nuclei. A) Watershed map. B) Ground truth for four overlapping nuclei. C) Ground truth for two overlapping nuclei

5.2.1.3 Performance of our automatic nuclei segmentation in the epidermis layer

a) *Feature Extraction*

There are two categories of image properties related to features that were used in my work, together with various combinations of these categories. The first category is based on the number of pixels belonging to a given region. In *Matlab*, this is implemented by the (Area) function, which extracts the binary object in the image and counts the number of its pixels. The second category is (MajorAxisLength) which was employed to find the length of the major axis for a fitting ellipse of the binary object.

b) *Classification*

We extracted five features from each segmented region (candidate nuclei): the Area and Major and Minor Axis of the best fitting ellipse, circularity, and aspect ratio. We then used these five features into a known classification scheme that train and test the performance of our nuclei counting scheme in the epidermis over our experimental dataset of mouse skin images. Herein, the Support Vector Machine (SVM) classifier was used for skin nuclei classification, because the SVM is the best machine learning technique to separate two different classes. In our case we used SVM to classify nuclei objects and non-nuclei objects.

5.2.1.4 Classification Results

To test the performance of our nuclei segmentation in the epidermis layer scheme, a number of experiments are conducted using our datasets of microscopic H&E images and adopting a number of training-testing protocols.

We selected 100 randomly of dataset H&E images, we selected 20 images for training and 80 images for testing.

In these experiments, the Support vector machine (SVM) method is used to obtain our optimal criteria to be used in our proposed nuclei segmentation scheme. The support vector machine (SVM) is one of the most widely used supervised classifiers which aims to find an optimal separating hyper-plane among different classes of a given n-dimensional dataset (training set). The optimal separating hyper-plane is the one that has a maximum distance to the nearest data samples (the so-called support vectors) in the training set. Such optimisation technique endeavours to maximise the margin between the hyper- plane and the support vectors, expecting a better classification accuracy.

A linear SVM classifier is used to classify the segmented objects in the epidermis layer of H&E images into nuclei and not nuclei objects, based on the five features explained above.

SVM is used to determine the best shape of each nuclei in the images based on two above filters. Score of the classification was found by *eq(5.7*) (Cristianini, Shawe-Taylor and others, 2000). Figure 5.16 shows the SVM hyper-plane, and the support vector example.

$$score = \sum_{i=0}^{n} a_i k(V_{i\,,} x_1) + b \qquad \textbf{\textit{(eq5.7)}}$$

Where $\boldsymbol{V}$ is the support vectors set, $\boldsymbol{n}$ is the size of $\boldsymbol{V}$, $\boldsymbol{a}$ is the alpha weight of *V*, *K* kernel function (a dot product <V_i, x_1> in this case) and *b* is the bias.

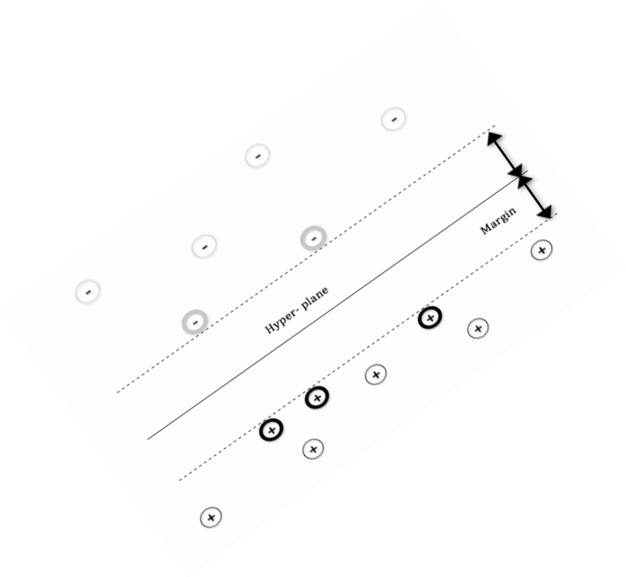


Figure 5.16 Example of an SVM classifier with an optimal hyper-plane between the two classes of support vectors. The grey vectors represent non-nuclei objects in the epidermis layer, and black vectors represent the nuclei objects in the images. The hyper-plane separate

Figure 5.17 shows the results of comparing the ground truth and the automatic segmentation before the improvement over 20 randomly-selected images where the blue bar represent the ground truth and the red bar represented the number of nuclei generated automatically. The x-axis is indicating the 20 images selected randomly and the y-axis represent numbers of nuclei segmentation in the epidermis layer.

For each candidate nuclei, we extracted the five features: the Area and Major and Minor Axis of the best fitting ellipse, circularity, and aspect ratio. The features are then fed into a liner SVM and an overall accuracy of 88.78% was achieved.

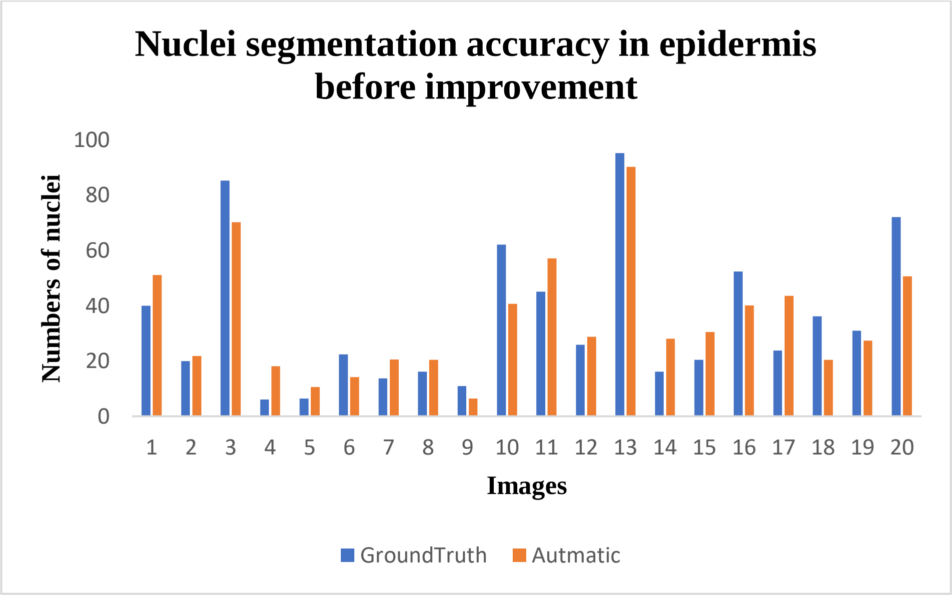


Figure 5.17 Nuclei segmentation in the epidermis layer: Comparison of automatic method to the ground truth

Figure 5.18, on the other hand, shows the accuracy of the SVM classification using the five features explained above after identifying and splitting the overlapping nuclei in the epidermis. An overall accuracy of 94.21% accuracy was achieved on the same 20 images. This improvement in results confirms the effectiveness of our scheme to segment overlapping nuclei in epidermis layer.

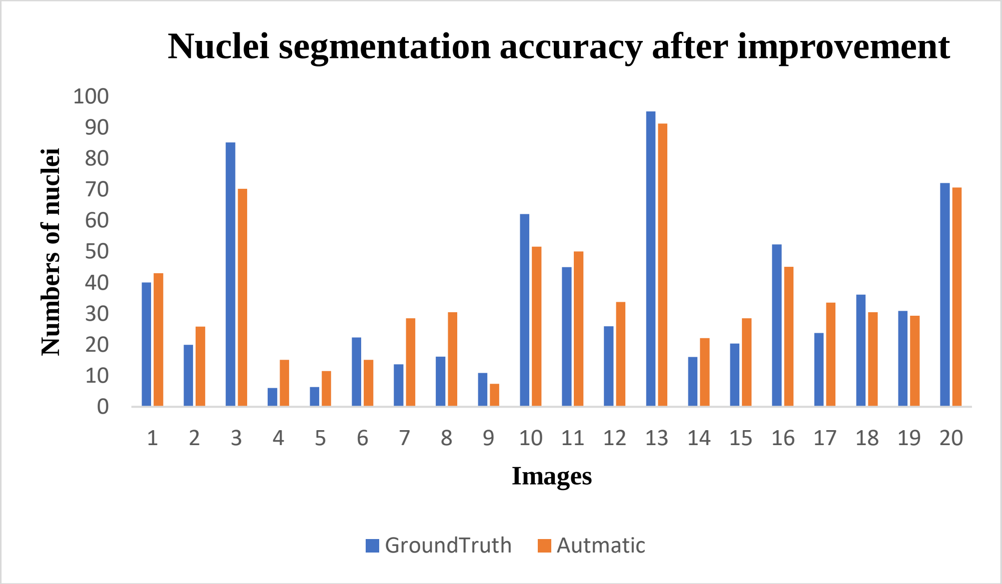


Figure 5.18 Overlapping nuclei segmentation in the epidermis layer after segmentation improvement: Comparison of automatic method to the ground truth

5.2.2 Nuclei Segmentation of Dermis Layer

In Chapter 4, the dermis layer was segmented, by applying a hybrid segmentation based on colour convolution and fuzzy clustering procedure. Our proposed nuclei segmentation in dermis layer, exploits the results of the Epidermis FCM threshold procedure (Figure 5.6C, displayed below as Figure 5.21A) because it does additionally detect the nuclei-like objects within the dermis layer. The algorithms work in two steps.

5.2.2.1 Detection of Nuclei-like objects in the Dermis Layer

Figure 5.19, show the block diagram of Nuclei segmentation of the dermis layer, which starts with the Dermis layer image (segmented in chapter 4, and displayed below as Figure 5.20C) as input. Hence, our algorithm first isolates the highlighted nuclei-like objects in the dermis layer by inverting the epidermis FCM threshold image (see Figure 5.20B) and multiplying with the dermis layer image (Figure 5.20C). Figure 5.20D, illustrated the result of this step for the sampled case. Finally, the nuclei-like objects are then segmented by multiplying the current image again with dermis layer image. Although, the output from this step (as illustrated in Figure 5.20F) detects the required objects, it includes other objects that need to be illuminated by a refinement procedure.

5.2.2.2 Refining Nuclei Segmentation of Dermis Layer

The segmented nuclei-like objects in the dermis layer needs to analysed to determine the genuine nuclei objects and illuminate the many unwanted objects in the output image from the last section. To perform this refinement step, we need to extract a suitable feature from the highlighted objects that clearly separate genuine nuclei-like objects from the non-nuclei objects (to be excluded). The area of the objects is an obvious easy to extract objects feature that could identify whether an object is noise or not. Several trial and error experiments on many images, revealed that the area of non-nuclei or noise was larger than 300. Hence, our refinement procedure first excludes any object that has an area of 300 pixels or more. Unfortunately, the output still includes non-nuclei objects, and circularity is another object features that was employed in the refinement step to further exclude some objects that are of appropriate size but not nuclei in the dermis layer. This choice is based on existing knowledge of the anatomy of mice skin, that the nuclei in the dermis are of circular shape, and not all objects in the dermis are nuclei, some of them were blood and lymph vessels as shown in Figure 5.21 (Wang *et al.*, 2014). The output from this step for our illustrated sample image is shown in Figure 5.20G

The whole steps of the proposed algorithm is displayed in the block diagram of Figure 5.19, while the various images in Figure 5.20 show the output at the various steps.

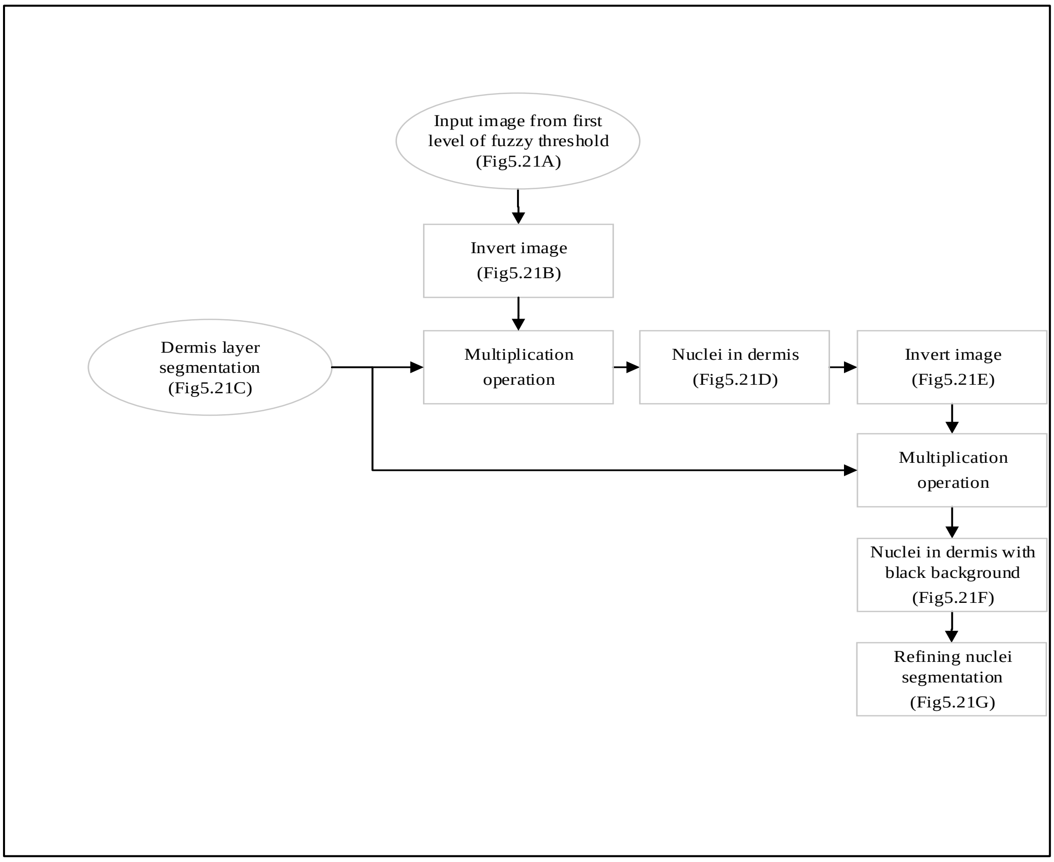


Figure 5.19 Block diagram of nuclei segmentation in the dermis layer

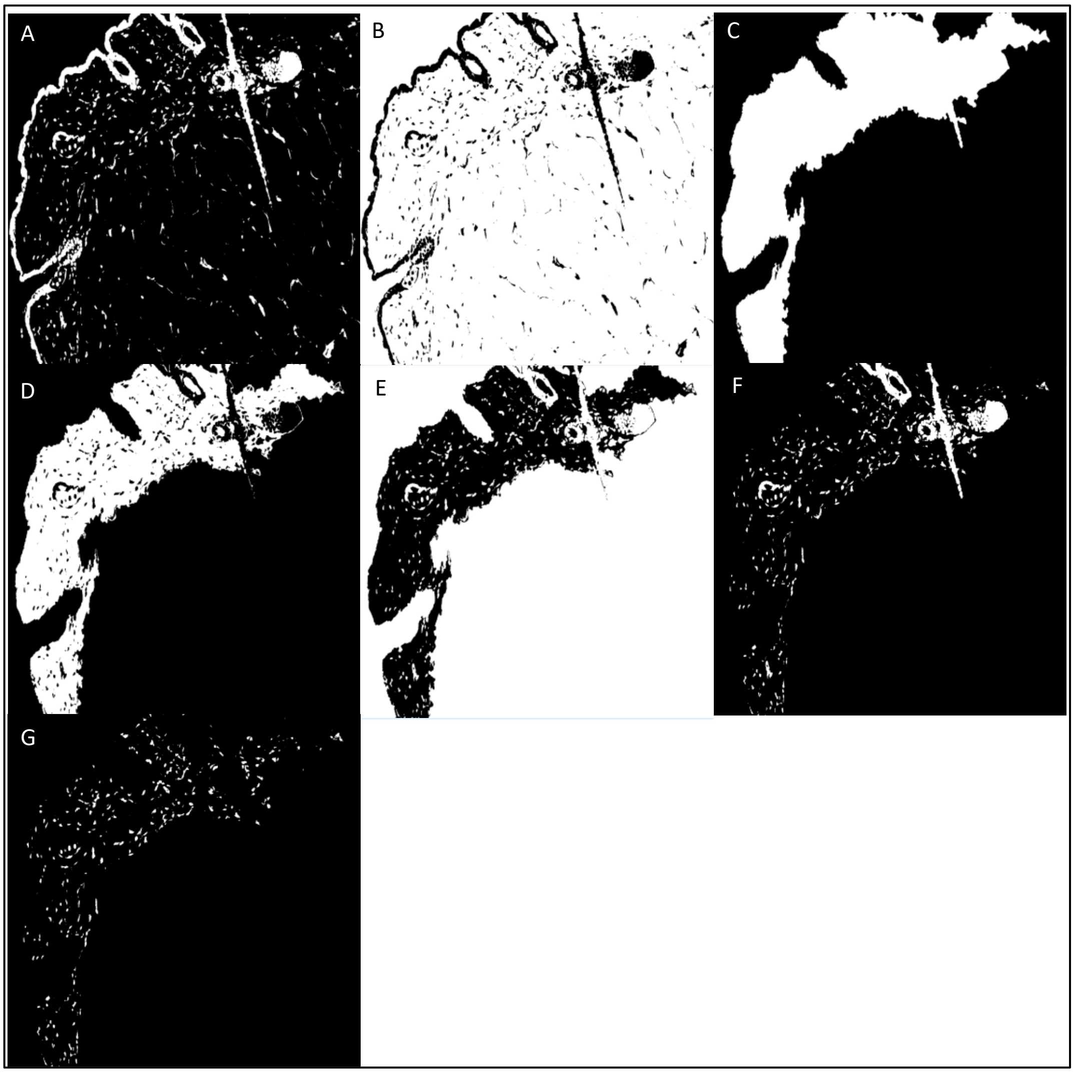


Figure 5.20 Nuclei segmentation in the dermis layer. A) Nuclei segmentation in all layers. B) Complement of image A. C) Dermis layer segmentation. D) Outcome of multiplication image C and image B. E) Nuclei segmentation with white background. F) Nuclei segmentation with black background multiplication image E and C. G) Improved nuclei segmentation by exclude the objects, which have area value 300 or more

Figures 5.21, below, displays the segmented nuclei-like objects on the RGB image of a sample skin tissues and point to a non-nuclei object as well as genuine nuclei objects. Figure 5.22, on the other hand show all the segmented genuine nuclei in the sampled skin tissue image.

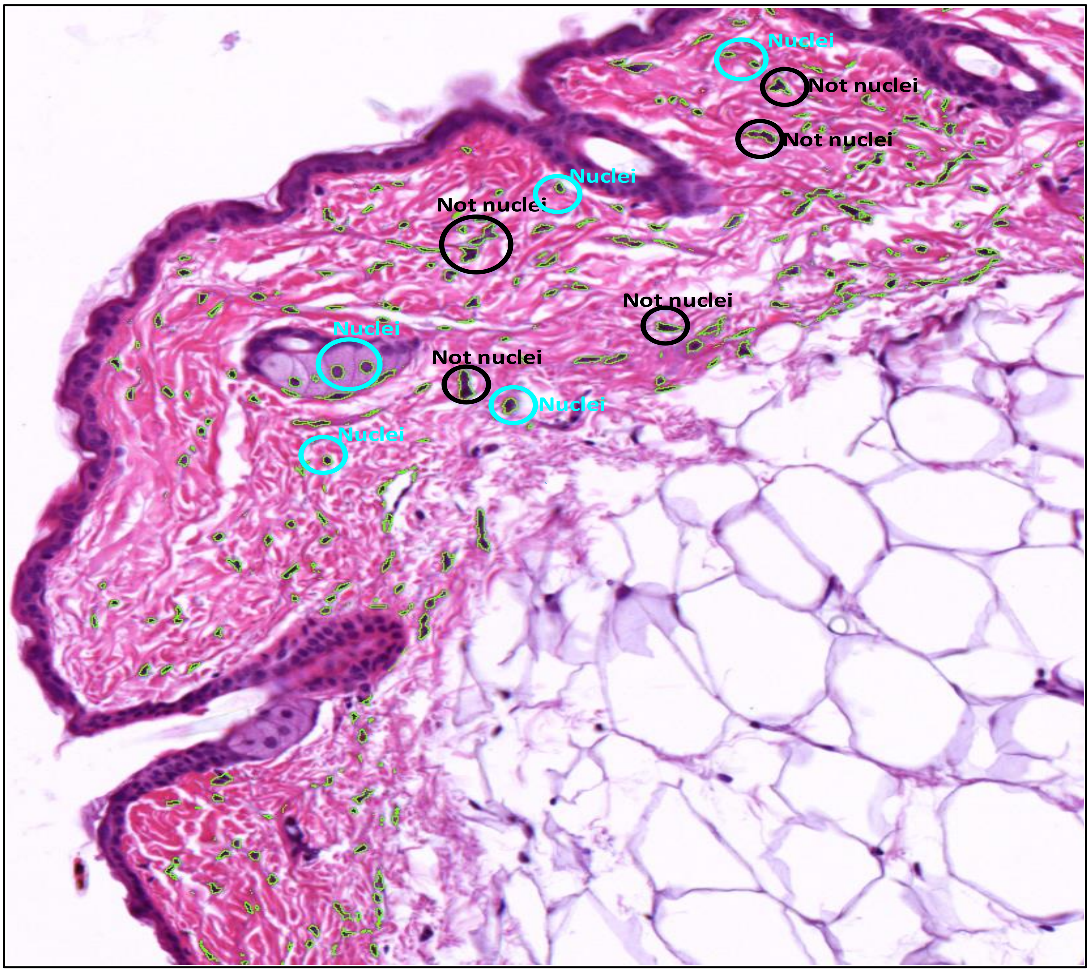


Figure 5.21 Nuclei and not nuclei segmentation

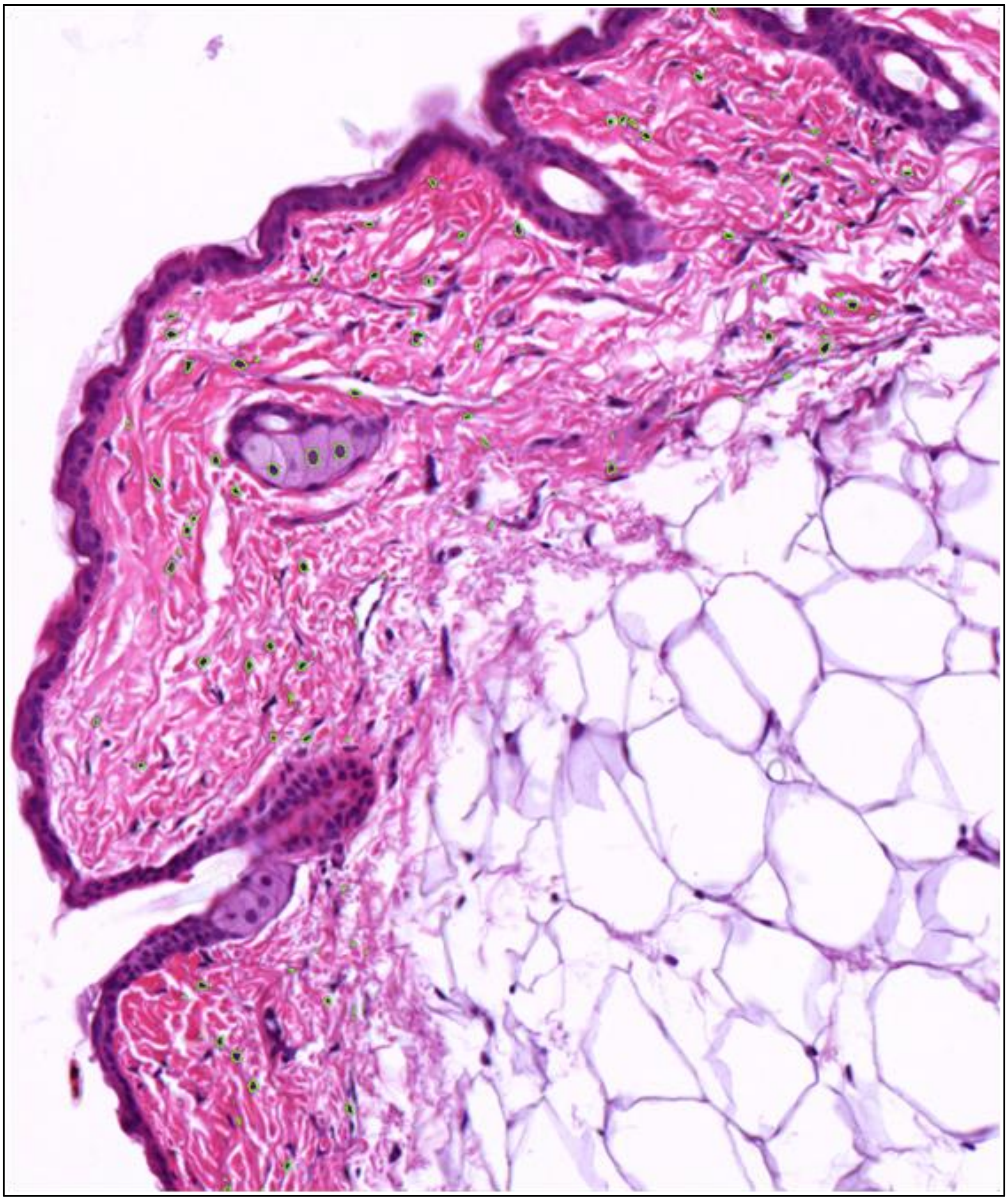


Figure 5.22 Nuclei segmentatoin in the dermis layer

5.2.2.3 Performance of the Nuclei Segmentation in the Dermis layer

Results of the SVM classifier were found, where the accuracy of the classification were 89.20% (before refining) and 95.02% (after refining) respectively. Figure 5.23 and Figure 5.24 show the difference in results between the automatic method and the ground truth. A blue bar is indicate to the ground truth, which we identify the nuclei segmentation manually in 20 images. A red bar is point to our automatic method, which is evaluated using SVM. We noted that an improvement in result as shown in Figure 5.24 due to apply our refining scheme to segment nuclei in the dermis layer.

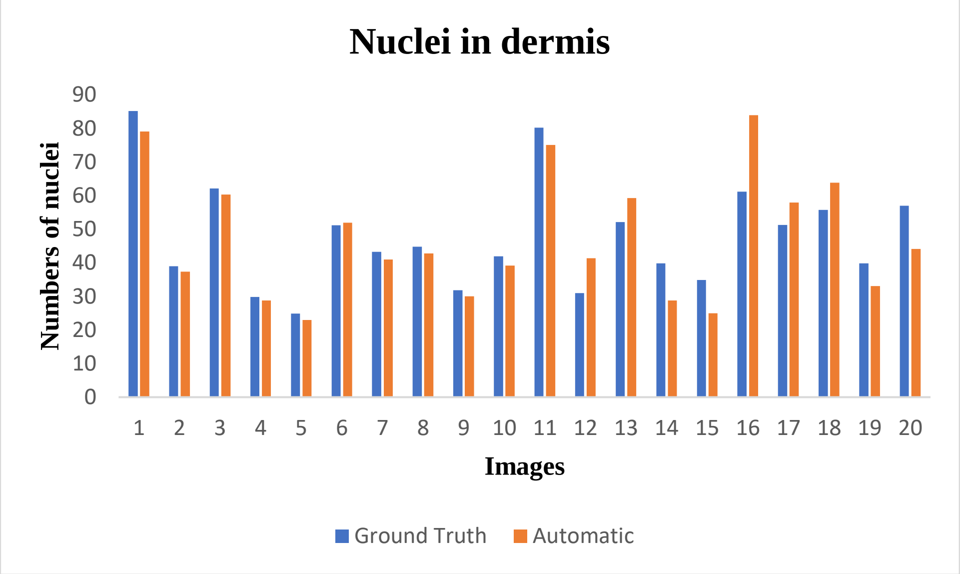


Figure 5.23 Nuclei segmentation in the dermis layer: Comparing automatic method and ground truth

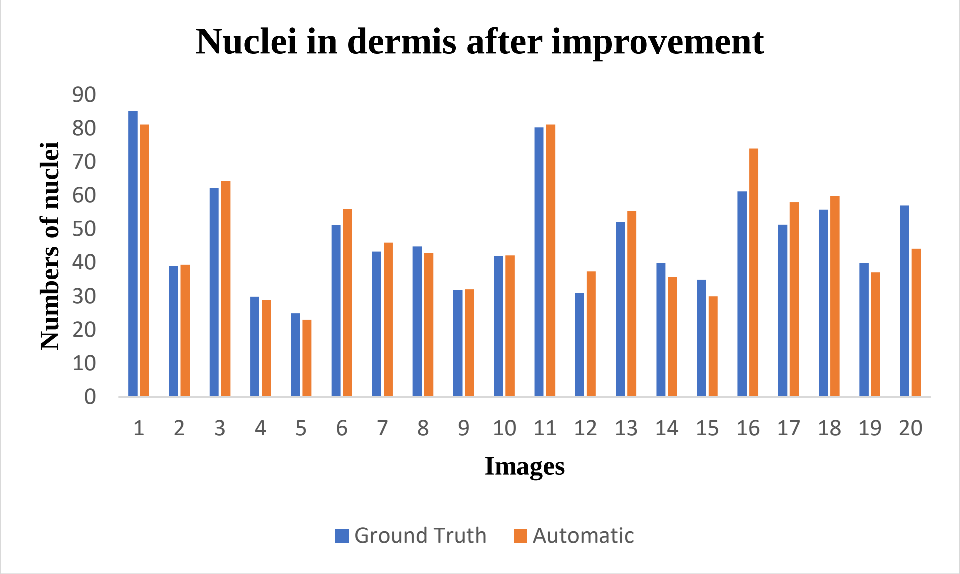


Figure 5.24 Nuclei segmentation in the dermis layer: Comparing automatic method and ground truth after segmentation improvement in dermis layer

The nuclei were then quantified by counting as the final goal and the nuclei measurements were extracted from a large number of images of mutant mice to be used in identifying a list of genes responsible for changes in the dermal nuclei.

5.3 Experiment Results and Analysis

In that last two sections, nuclei segmentation procedures, we used the supervised machine learning SVM classifier to test their performances based on certain extracted features from the segmented nuclei. We also, compared the results of using our automatic schemes with those obtained manually for a sample of 20 images. In this section, we present the results of two sets of experiments. The first set is centred on comparing the automatic measurements to those obtained manually while the second is dedicated to identifying genes responsible for changes in nuclei numbers in the dermis and epidermis layers. Due to the difficulty of manually labelling all the images in our dataset of mouse skin tissue images, we selected randomly a sample of 100 images that were annotated by an expert to be used as a ground truth.

5.3.1 Comparing Manual vs. Automatic Measurements

First of all, the number of nuclei in the two layers (epidermis and dermis) were calculated by applying the two proposal methods for 100 randomly selected images. The manual counting method was carried out by marking and counting the nuclei in the same set of 100 images by a domain expert. In the first evaluation, we check the correlation between the manual and automatic measurements using R^2 and ARL schemes (see Section 3.4.2.1 for further details). Figures 5.25A and B show the scatter plot for the manual versus automatic measurements for nuclei counting in two layers dermis and epidermis. The figures demonstrate close linear correlation between the manual and the automatic. However, the count of nuclei in the dermis layer still shows more consistency than the nuclei in the epidermis layer due to the underlying difficulty in identifying the nuclei in epidermis as explained earlier.

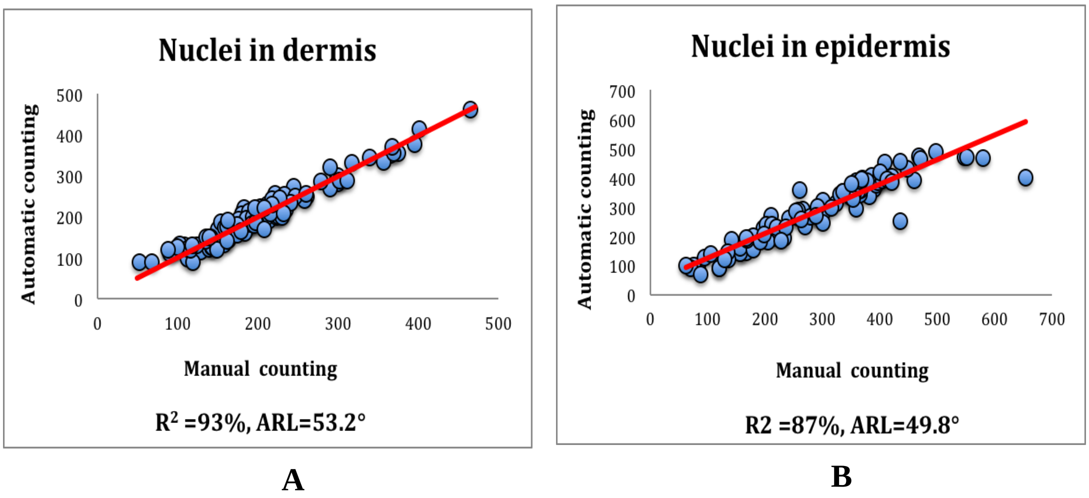


Figure 5.25 Correlation of manual versus automatic counting. A) Nuclei counting in the dermis layer. B) Nuclei counting in the epidermis layer

To further analyse the effectiveness of the proposed solutions, we compare the manual versus the automatic of the nuclei counting for dermis and epidermis, Table 5.1 shows the comparison results using four parameters: namely R^2, ARL, MAE, and the standard deviation (σ) of absolute error between the ground truth and the automatic measurements (see Section 3.4.2 for further details).

Table 5.1 Manual versus automatic nuclei counting in dermis and epidermis layers

Measure	R^2	ARL	MAE	σ
Nuclei counting in dermis layer	0.93	53.2°	0.135	3.53
Nuclei counting in epidermis layer	0.87	49.8°	0.28	2.82

The table shows noticeable difference between the manual and the automatic measurements especially for the nuclei counting in dermis and epidermis layers, which shows the counting in dermis is better than in epidermis. This is probably due to the observed of effect of staining in showing the nuclei in the epidermis layer mostly overlapping with the surround tissue. While the nuclei in the dermis are mostly prominent within their surrounding tissue which explain the need for two separate methods. However, nuclei counting errors in the dermis layer is due to the fact that not all objects are nuclei.

To further assess the correlation between the manual and automatic measurements, we examined their closeness using the ***Bland Altman analysis*** (BA) (see Section 3.4.2.2).

Table 5.2 Manual versus automatic for nuclei counting in dermis and epidermis layers

Measure	Upper LoA	Lower LoA
Nuclei counting in dermis layer	36.91	-37.93
Nuclei counting in epidermis layer	92.42	-79.44

Table 5.2 shows the lower and upper LoA for all proposed measurements. Figure 5.26 and Figure 5.27 shows the BA analysis scatter plot for the final nuclei counting in dermis and the final counting in epidermis respectively, which is display closeness between automatic method and ground truth (see section 3.4.2 for further details) by using 100 images randomly.

Again, the results in Table 5.2 confirm the conclusion of Table 5.1 and show that nuclei counting in dermis are better accuracy from nuclei counting in epidermis, the reason behind that is LoA range for nuclei segmentation in the dermis layer, which is start from -37.93 to 36.91 and most of the points located inside the range as shown in Figure 5.26. While the LoA for the nuclei segmentation in the epidermis layer is bigger with range between -79.44 to 92.42 with some points outside the range as shown in Figure 5.27.

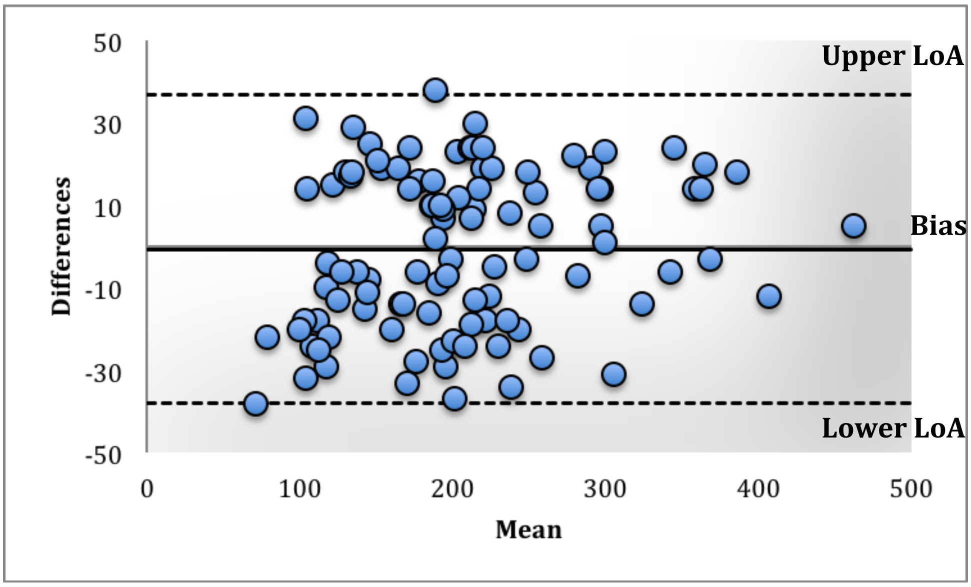


Figure 5.26 Bland Altman analysis for manual versus automatic for nuclei counting in the dermis layer

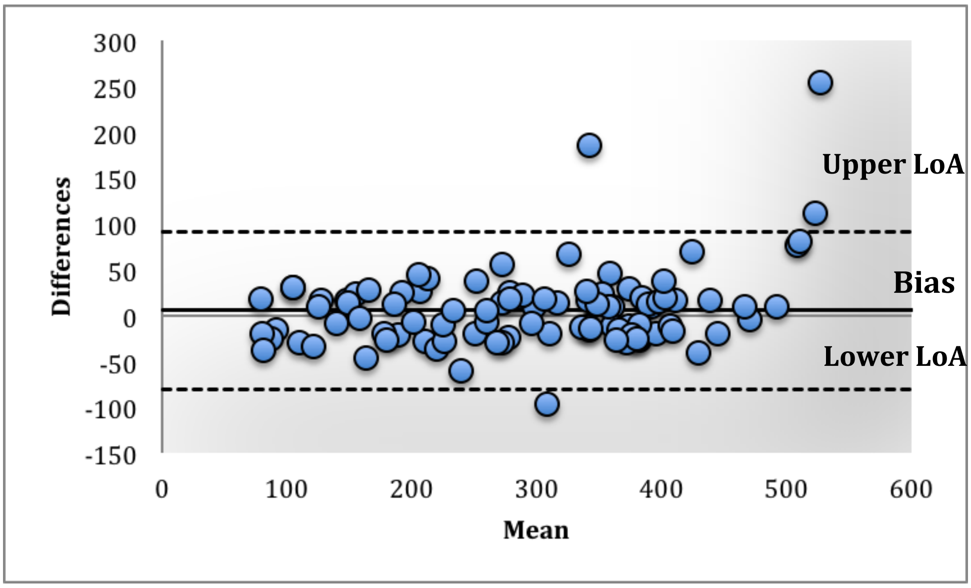


Figure 5.27 Bland Altman analysis for manual versus automatic for nuclei counting in the epidermis layer

5.3.2 Genes Identification Experiments

Having demonstrated high accuracy rates of the two algorithms for segmented nuclei in the epidermis and dermis skin layers, in this section we report the results of our experiments to identify the genetic basis of abnormalities in the nuclei in two layers dermis and epidermis i.e. what are the genes that may lead to an increase or decrease the nuclei counting? These experiments will be apply the proposed methods on the whole dataset of the 7000 H&E skin images (described in Section 3, Chapter 3). To do this, first needs to establish what is normal and what is not when it comes to nuclei counting of mice skin. This is typically done based on the RR approach to identify the Lower Bound (LB) and the Upper Bound (UB) of the nuclei counting obtained from WT animals (see Chapter 3 Section 4.1 for further details). For illustration purposes, Figures 5.28 show the distributions of the nuclei counting obtained from 500 WT images. The LB and UB of the RR are the vertical black lines on the two sides on the charts. Next, two main steps were achieved to identify the interesting genes step1 and step2

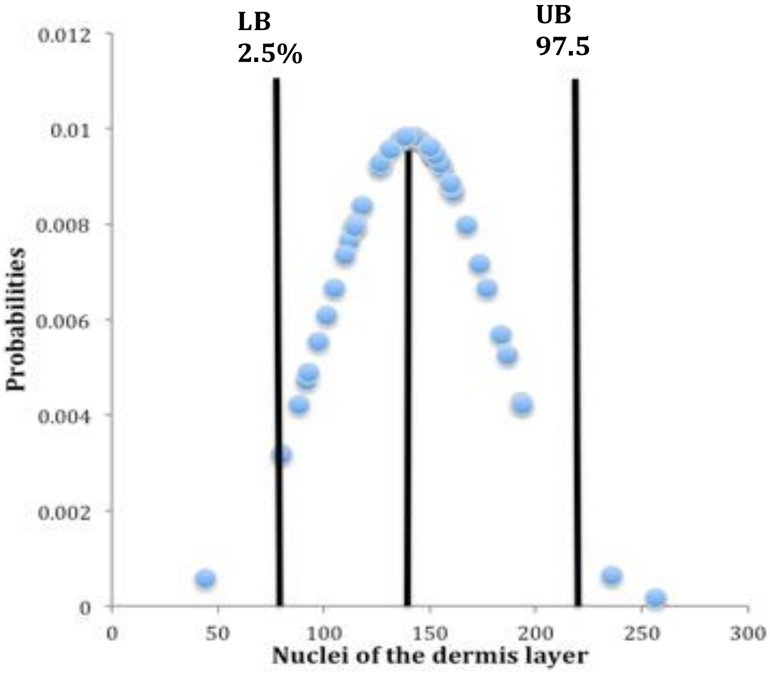


Figure 5.28 Normal distribution of nuclei in the dermis layer for WT population with LB and UB obtained from RR

Step 1:

Nuclei number of every mutant mouse were compared against the relevant LB and UB to establish whether the measurement falls within or outside the RR. Similar to WT images, Figure 5.29 shows the distribution of the 261 mutant mice. The vertical black lines represent the LB and UB of the RR (74 to 220), obtained from the nuclei number in dermis of WT (normal) cases.

Out of 6500 mutant mice images analysed, 5850 (90.03%) were found to be within the RR established by the WT measurements, and 650 images (9.6%) of mutants fell outside the RR for nuclei in dermis layer. However, it is important to highlight that not all samples that fall outside the RR are necessarily true hits (abnormal cases). Arguably the above number includes many false cases due to two main reasons. First is related to any incorrect nuclei detection, which leads to incorrect nuclei counting. Secondly, although correctly identified abnormal cases may belong to a specific gene for which the total number of cases of abnormal nuclei number is less than the hit rate threshold (see Section 3.4.1 for further details).

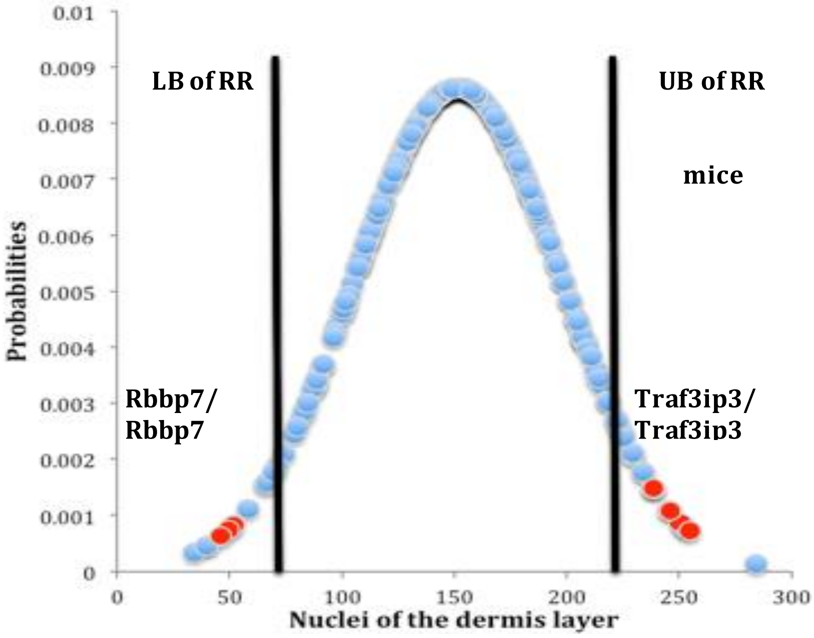


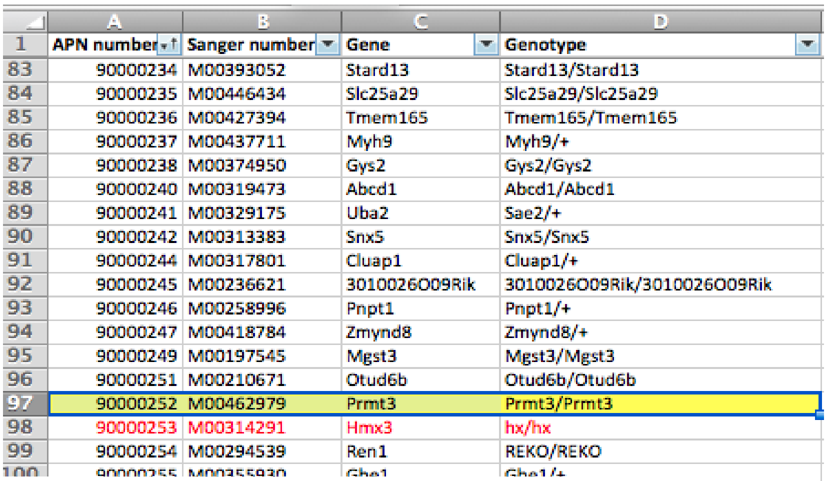
Figure 5.29 Normal distribution of nuclei in dermis layer for mutant population with RR bound

Step 2:

To filter out the false cases, we calculated the so-called hit ratio of each mutant line (see Section 3.4.1 for details). Out of the 650 outlier cases, only 56 cases were considered as true hits (i.e. hit ratio > 60%) as shown in Table 5.3 where the "-" mark is used to indicate that the ratio is below 60%.

The following steps identified interesting genes:

1- Datasets collection from Sanger institute, where each animal has unique Access Point Name (APN) and sanger number with genes and genotypes as shown in Figure 5.30.



	A	B	C	D
1	APN number	Sanger number	Gene	Genotype
83	90000234	M00393052	Stard13	Stard13/Stard13
84	90000235	M00446434	Slc25a29	Slc25a29/Slc25a29
85	90000236	M00427394	Tmem165	Tmem165/Tmem165
86	90000237	M00437711	Myh9	Myh9/+
87	90000238	M00374950	Gys2	Gys2/Gys2
88	90000240	M00319473	Abcd1	Abcd1/Abcd1
89	90000241	M00329175	Uba2	Sae2/+
90	90000242	M00313383	Snx5	Snx5/Snx5
91	90000244	M00317801	Cluap1	Cluap1/+
92	90000245	M00236621	3010026O09Rik	3010026O09Rik/3010026O09Rik
93	90000246	M00258996	Pnpt1	Pnpt1/+
94	90000247	M00418784	Zmynd8	Zmynd8/+
95	90000249	M00197545	Mgst3	Mgst3/Mgst3
96	90000251	M00210671	Otud6b	Otud6b/Otud6b
97	90000252	M00462979	Prmt3	Prmt3/Prmt3
98	90000253	M00314291	Hmx3	hx/hx
99	90000254	M00294539	Ren1	REKO/REKO

Figure 5.30 Sanger institute data sets

2- Images collection by Sanger institute which consists of each slice of image (a group of images of the same animal) and with type of slices such wild type (normal cases) or mutant (abnormal cases) for more details see (Chapter3 in Section3).

3- We applied our scheme on the data sets to count nuclei per image and analyse results i.e. determined the distribution data to identify LB and UB for wild type images (see Chapter3 in Section3). Then it is apply the LB and UB to the mutant type (step 1), to locate images if they are outside of RR (i.e. larger than UB or smaller than LB), we considered as interesting cases or inside RR (i.e. between UB and LB), we considered as normal cases as shown in Figure 5.31.

4- Determine hit ratio i.e. count a true cases and ignore false cases, by calculating the hit ratio, if it is larger than 60% (for example 5 out 7 images inside RR) we considered as true case, otherwise we considered false case (step2). Based on hit ratio we identify the interesting genes as final objective i.e. match our results (true cases) with Sanger institute data sets to identify the genotype.

Table 5.3 Possible gene hits in the dermis layer

Name of allele	Nuclei in the dermis layer	
Prmt3/Prmt3	>UB	83%
Traf3ip3/Traf3ip3	>UB	100%
Ppp3ca/Ppp3ca	>UB	73%
Psat1/+	Normal	-
Rhot2/Rhot2	Normal	-
Amotl1/Amotl1	Normal	-
Clk1/Clk1	<LB	95%
Ido1/Ido1	Normal	-
Cntfr/+	<LB	100%
Amfr/+	<LB	85%
Rbbp7/Rbbp7	<LB	86%
S100b/S100b	Normal	-
Lpar5/Lpar5	<LB	86%

The table above shows a list of 8 genes that could be potentially responsible for abnormalities in the nuclei in dermis layer, 3 genes show increasing in the nuclei and 5 genes show decrease in the nuclei in the dermis.

To large extent this confirms the effectiveness of our developed method in identifying the list of genes responsible for changes in the nuclei number in the dermis layer. Interestingly, the table shows a possible novel gene called ‘Traf3ip3/Traf3ip3’, which is currently shows that the knockout this gene causes increase in nuclei, which cause the tumour, which is melanoma cells (Nasarre *et al.*, 2018).

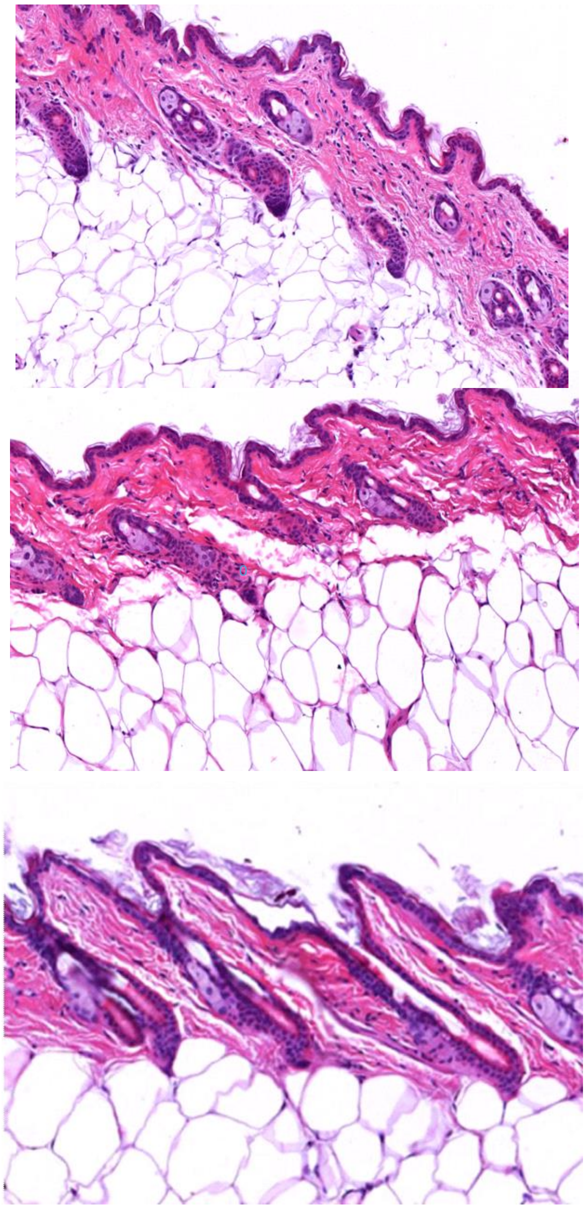
To identify the interesting genes in the epidermis layer, we utilize the same approaches in Section 5.5.2 to show the effective genes as illustrated in Table 5.4.

Table 5.4 Possible gene hits in the epidermis layer

Name of allele	Nuclei in the epidermis layer	
	Phenotype	Hit Ratio
Prmt3/Prmt3	>UB	86%
Amfr/Amfr	>UB	100%
Dusp3/Dusp3	>UB	86%
Abcd1/Abcd1	>UB	70%
Pik3cb/Pik3cb	>UB	94%
Pabpc4/Pabpc4	>UB	89%
Coq9/Coq9	>UB	91%
Kng2/Kng2	>UB	100%
Rcor2/Rcor2	>UB	94%
Pld5/Pld5	Normal	-
Suv420h1/Suv420h1	Normal	-
Psat1/+	Normal	-
Myo5a/Myo5a	<LB	100%
wi/wi	<LB	74%
Cish/Cish	<LB	81%

Where the above table 5.4 shows 12 genes with abnormal cases in the epidermis layer, 9 genes function to increase the nuclei number and 3 genes function to decrease the nuclei.

From two tables (table 5.3 and table 5.4) there is one common gene call "Prmt3/Prmt3" which show clearly in the both layers dermis and epidermis causes increasing the nuclei as shown in Figure 5.31.



Prmt3/Prmt3

(+/+)

Lpar5/Lpar5

Figure 5.31 Images of genotypes with that influence number of nuclei compared to wt

5.4 Conclusion

Automatic methods for nuclei segmenting and counting have developed for the epidermis layer and dermis layer in mice skin. Building on the work in chapter 4, the output of the our hybrid method procedures provided the obvious starting input of the algorithms in that it provided effective segmentation of the two layers dermis and epidermis containing the nuclei objects. However, we applying appropriate thresholding to segment the nuclei in two layers (epidermis and dermis), where the output still includes many non-nuclei objects besides the problem of overlapping objects. We incorporate two subsequent filtering using certain known nuclei features (circularity, aspect ratio, major axis, perimeter) to effectively automatically identify overlapping nuclei in the epidermis layer, which were then separated using the watershed transformation. Knowledge about the circularity shape of nuclei together with their expected sizes in mouse skin tissue was also essential in eliminating the largest proportion of non-nuclei objects in the dermis layer. The performance of the various steps and procedures were evaluated using the SVM classifier. The result of nuclei segmentation in the epidermis layer, assessed by SVM, achieved an overall accuracy of 88.78%. The corresponding accuracy for the identification of overlapping was 94.21%. Results of the SVM classifier on the performance of our second algorithm of segmenting and quantifying the nuclei in the dermis layer were equally significant at 89.20% and after refining was 95.02%. In all cases the true acceptance and true rejection rates were close to each other and to the overall accuracy, demonstrating the significance of the contribution of this chapter.

Subsequently, the proposed method helped in isolating a potentially interesting (knockout genes), further indicating the potential of the proposed methods. The evaluation of our performances shows success and viability for incorporation into lab systems based on domain expert. The development of effective, non-invasive and wholly unsupervised techniques would facilitate high-throughput analysis in cutaneous research, with potential applications for screening drugs.

The next two chapters will be devoted to developing equally effective automatic algorithms to quantify the other skin layer factors associated with skin abnormalities.

Chapter 6 : Hair Follicle Orientation of Mice Skin

Having developed algorithms to count nuclei in the epidermis and dermis skin layers and investigated their benefit to identifying genes that may cause skin abnormality, in this chapter, we investigate factors relating to another skin object for the same end. Hair follicles cover most of the surface of mammalian bodies, and it has been shown that the orientation of follicles is associated with some genetic disorders such as hypotrichosis simplex or monilithrex (Sundberg, 1994). The Mouse Genetics Project (MGP) provides a valuable resources to identify the exact genes responsible for changes in the orientation of hair follicles (Chang, Philip M Smallwood, *et al.*, 2016).

The main aim of this analysis is to identify and quantify the hair follicles in the epidermis layer using automated high throughput analysis. We shall present an algorithm to quantify the orientation of hair follicles in the epidermis layer, and assess changes in their orientation in relation to genetic determinants of skin organisation i.e. genetic causes of relevance skin disease such as hypotrichosis simplex or monilithrex (Sundberg, 1994). This chapter has two main contributions:

1. Developing an automatic method compute the orientation of hair follicle with respect to a fixed alignment of the epidermis layer.
2. Design an effective segmentation and quantification of hair follicles’ orientation in the epidermis of wild type (WT) and mutant type mouse skin images:.

The rest of this chapter is organised as follows. In Section 6.1 the problem statement of quantifying hair follicle orientation is described, highlighting the main challenges. In Section 6.2 we present our automatic method and explain the adopted image processing and analysis techniques in details. Section 6.3, report and analyse the experimental results for identifying genes associated with changes in hair follicles’ orientation in the epidermis of mice skin. Finally, in Section 6.4, the chapter is concluded by discussing and concluding the effectiveness and efficiency of the proposed solution.

6.1 Problem Statement

The orientation of the hair follicles with respects to the epidermis is perceived to be an indicator of the presence/absence of skin abnormalities for a variety of reasons. Where the orientation of hair follicle has steady in the normal cases, while in abnormal causes hair follicle disorientation in the skin (Wang *et al.,* 2016). Figure 6.1, illustrates the differences in orientations of hair follicles in healthy and abnormal skin conditions.

The main challenge in the chapter “How to automatically measure the angle of orientation of hair follicle with respect to a fixed alignment of the epidermis layer and accurately quantify the orientation of follicles per image from an H&E image of a mouse skin tissue”.

Manual quantification of hair follicle is time consuming and tedious. This is due the large number of hair follicles that need to be processed in each image, and there can be a huge number of H&E images that are continuously generated by the problem domain (i.e. MGP) and other skin abnormality related research objectives.

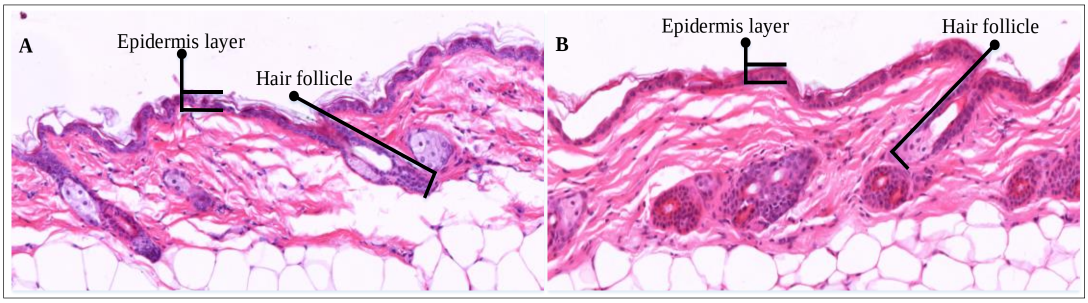


Figure 6.1 Placement of the hair follicle in the normal and abnormal image. A) Normal image. B) Abnormal image

Before we discuss our automatic proposal method of quantifying the orientation of hair follicles in the epidermis, we must understand the way this task is manually conducted. We first note that our experimental 20X magnification microscopic skin images clearly highlight the structure of the various layers of mouse skin, namely the epidermis, dermis and subcutaneous layers. Hence, the human lab researcher first needs to select a suitable way on measuring the hair follicle orientation manually. Traditionally, quantify hair follicles is done by measuring, under the microscope, the angles between the hair follicle (HF) and epidermis layer (van der Veen *et al.*, 1999), as shown in Figure 6.2. Because it was possible to identify the HF in skin layers, and it is a particularly attractive goal for 20X magnification screens because of their clear dynamic nature and their easy accessibility (Awgulewitsch, 2003; DiTommaso *et al.*, 2014).

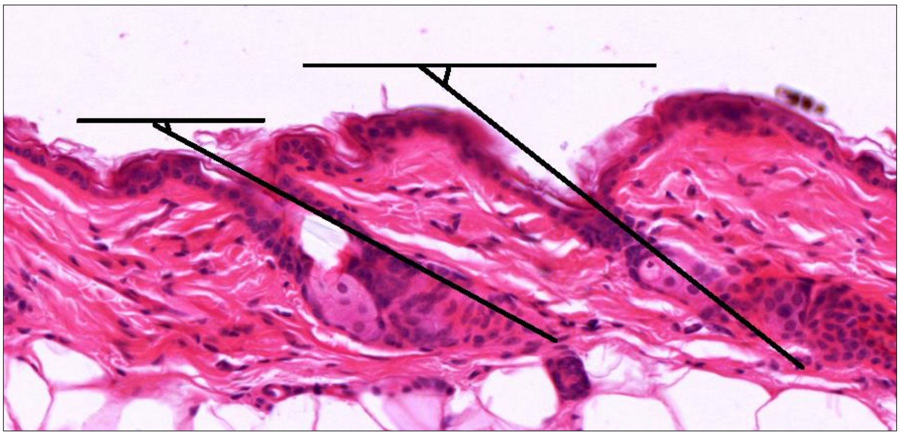


Figure 6.2 Quantify the hair follicle manually

To develop an efficient automatic scheme of high accuracy, however, we first need segment the hair follicles in the epidermis and then develop a criterion for determining orientation of the segmented objects with respect to the surrounding epidermis layer. Automatic segmentation and quantification of the hair follicle from H&E images with 20X magnification is made more complex by variation in their positions within the epidermis. We can identify the following challenges:

First: The source images in our study are full multi-layers skin images and hence accurate segmentation of ROI, i.e. the hair follicle in the epidermis areas, is an essential and critical pre-processing step. The use of the automatic scheme of Chapter 4, for epidermis layer segmentation, will result in narrowing the research space. However, colour intensity of hair follicles is close to surround tissue and their separation requires a great deal of concentration of the mind and a good level of experience, see Figure 6.1.

Second: There is a significant variation in the alignments of the epidermis across the different input images. Figure 6.3, displays two images with different epidermis alignment. This makes hair follicle orientation difficult to compare between images, unless we realign the epidermis in the input image to an agreed direction.

Third: A hair follicle is a curved shape, quantifying its orientation (i.e. its angle to the aligned epidermis) is unstable unless we fix a criterion for computing hair follicle orientation angle independently of the follicle curvature through its length, to be used for all follicles within each and all images.

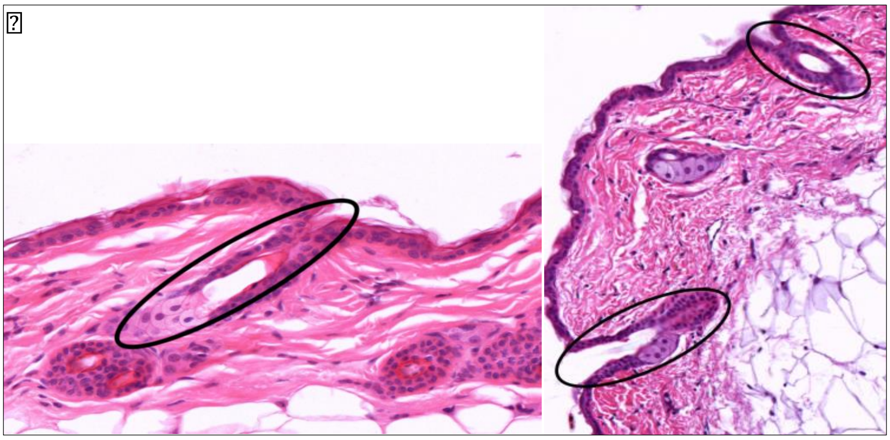


Figure 6.3 Quantify the hair follicle manually

6.2 The proposed Method for Quantifying Hair Follicle Orientation

In this section, we propose a system to detect morphological features of the hair follicle in the epidermis layers of the skin. We introduce image-processing techniques to segment the hair follicle and quantify the orientation of follicle. Our adopted methods for quantifying the hair follicle consists of three main parts. Naturally, first step will be the use of the algorithms of chapter four to partition regions of interest in histopathological skin images into three layers (i.e. epidermis, dermis and subcutis). This would be followed by an algorithm that segment hair follicles from the region of the epidermis layer, and align the epidermis layer in a normalised direction from left to right sides of the image, the top and bottom. Thirdly, we develop and test the performance of algorithms for the automatic quantification of biologically relevant parameters associated with orientation of hair follicle in epidermis layers in images generated by the MGP.

Figure 6.4 illustrates the main stages of the proposal method. The process at each stage will be further explained in the following subsections, and for the sake of completeness the first subsection will remind the readers of the epidermis segmentation method.

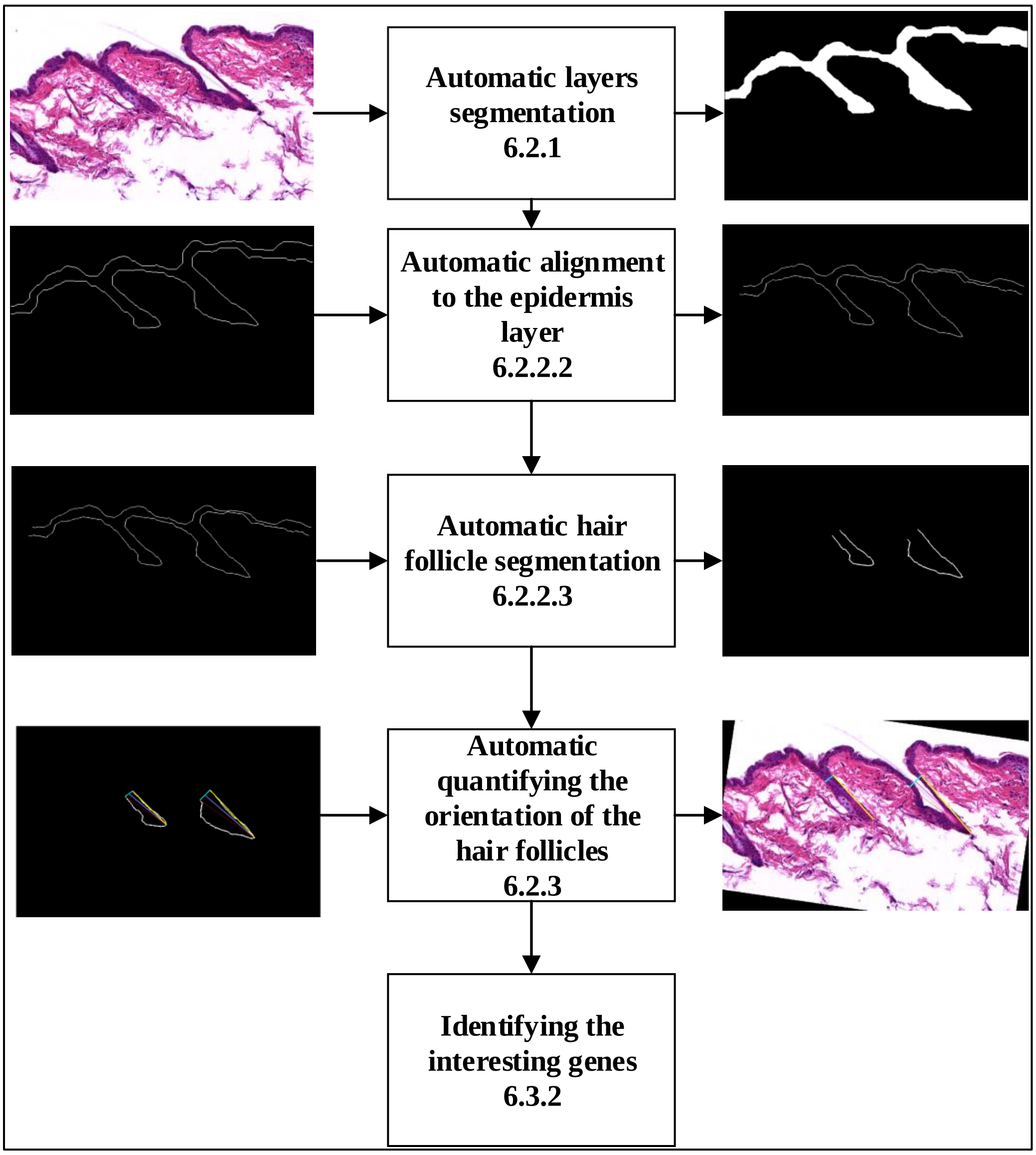


Figure 6.4 The main stages of the proposed method

6.2.1 Epidermis Layers segmentation

Within Chapter 4, we proposed and tested a set of algorithm that use the fuzzy c-mean clustering for automatic partitioning of H&E stained mouse skin images into different skin layers individually. This step will be implemented prior to enabling the segmentation of the hair follicles and other associated objects in the epidermis layer.

6.2.2 Hair Follicle Segmentation

The developed scheme to segment the hair follicles consist of three procedure to be implemented sequentially: (1) isolating two bounding curves of the outer thin sublayer of the epidermis layer, (2) align these two curves, and (3) segment the hair follicles. We shall now describe these procedures.

6.2.2.1 Evaluation of the Top and Bottom Curve Section of the Epidermis Layer

The border of the epidermis binary mask was found by subtracting two epidermis masks: the original mask and a second mask derived from the first by one pixel erosion morphology thereafter to be called the erosion mask. The very outer curve bounding this sublayer will be referred to as the top curve, while the bottom curve is the curve that separate the sublayer with the interior of the remaining part of the epidermis. The shape of the bottom curve is associated with the hair follicles in the epidermis layer (Huang, 2008; Koch *et al.*, 2011; Jiang *et al.*, 2012; Yamasaki *et al.*, 2012; Wu *et al.*, 2014; Praveen and Agrawal, 2015). The positions of the hair follicles by appearance of deep pocket like sections along this curve. The C, D and F images in Figure 6.5, illustrate the steps of this procedure.

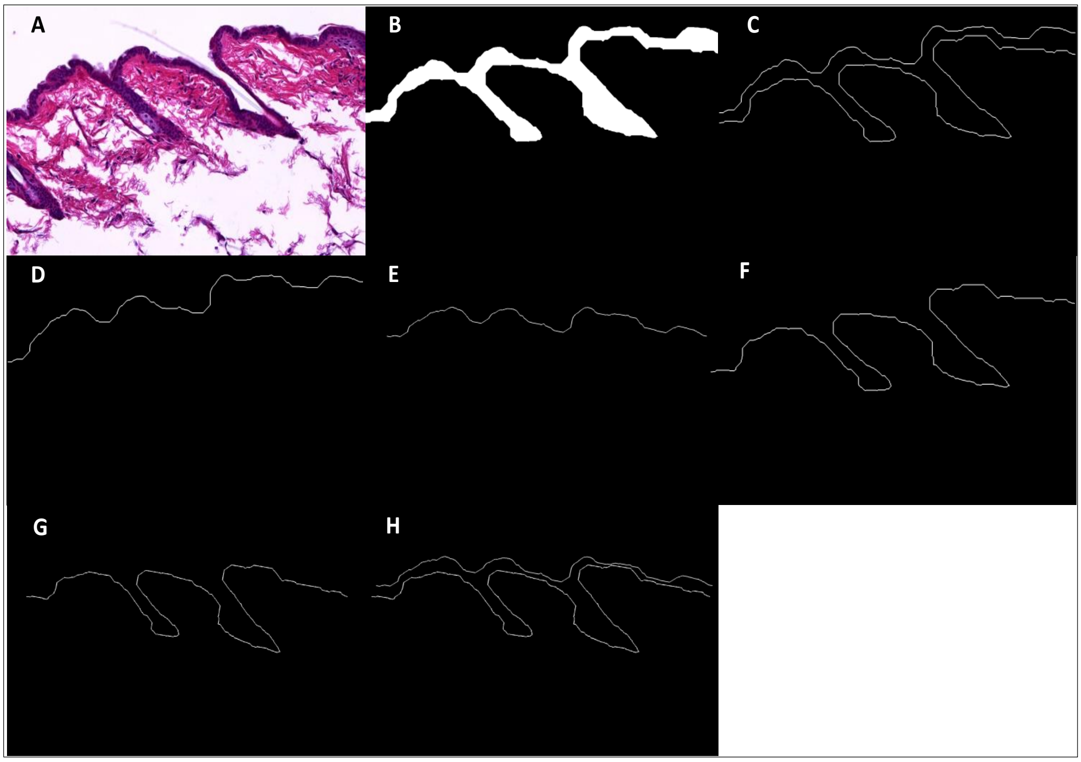


Figure 6.5 Automatic methods to align epidermis layer in H&E images. A) Original image. B) Segmented Epidermis layer. C) Epidermis layer border. D) Top line of epidermis. E) Align the top curve by using slop of first point and last point in the object. F) Bottom object of epidermis layer. G) Align the bottom object by using slop of first point and last point in the object. H) Addition operation for top object and bottom to gather them as the epidermis layer

6.2.2.2 Alignment of the Top and Bottom Curve Section of the Epidermis Layer

Pre-processing to improve the quantitative analysis was required prior to the next step of hair follicle segmentation, in order to make the orientation of follicle comparable for all images. This is mainly aimed to realign the top and bottom curves obtained above to a normalised position, in order to enable effective comparisons of images in both types (wild type and mutant). It necessary to align all images and to develop a reasonable analytical result (Hughes *et al.*, 2013; Zhang *et al.*, 2013; Mustra, Grgic and Zovko-Cihlar, 2014). The procedure first identifies the first point and the last point for both the top and bottom curves (Figure 6.6). The pixel that has the minimum x-coordinate value defines the first point in the top curve and the last point is represented by the pixel of the maximum x-coordinate value. A line is then drawn from the first point to the last point of the top curve. Following this, the slope of the curve was determined by using the first and last points for each curve, and then the angles were identified by the slope according to the x-axis. Finally, we rotate the image by angle value to align the image as displayed in Figure 6.5 (E, G). Similarly, we determine the first and last point of the bottom curve and rotated similarly. Subsequently, the alignment for the top curve and the bottom curve was added to the same image, i.e. the new image has two aligned curves, as shown in Figure 6.5H. Furthermore, the aligned image was transferred to a larger image matrix, to ensure that hair follicles located at the edge of the image were not lost. In addition, the top curve and the bottom curve together in the same image are represented by the epidermis layer in mice skin. The following block diagram shows the whole processes to align the epidermis layer in Figure 6.7.

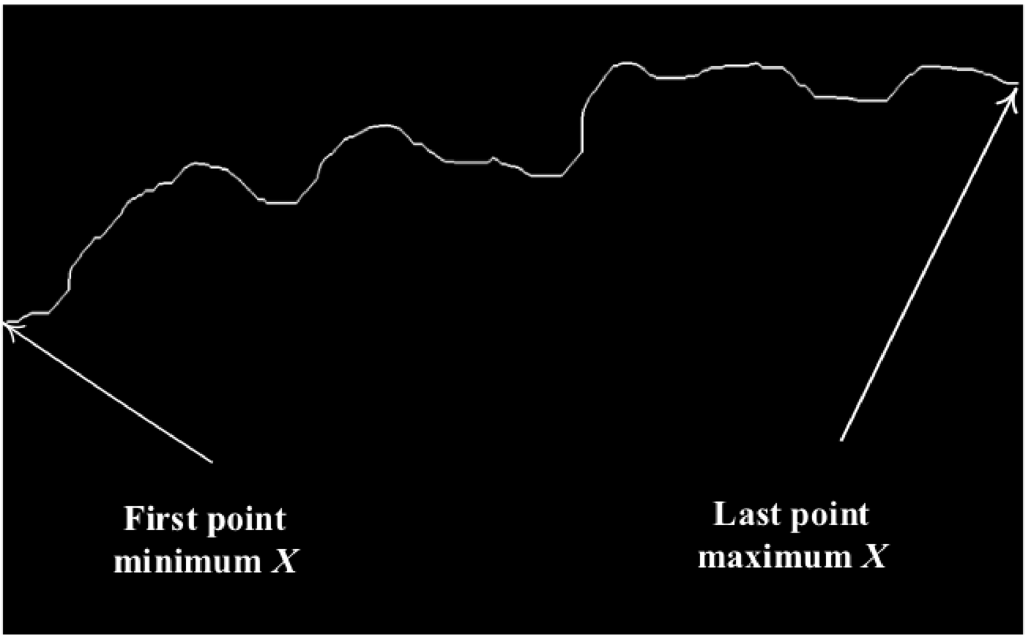


Figure 6.6 Identifying first point and last point in the curve

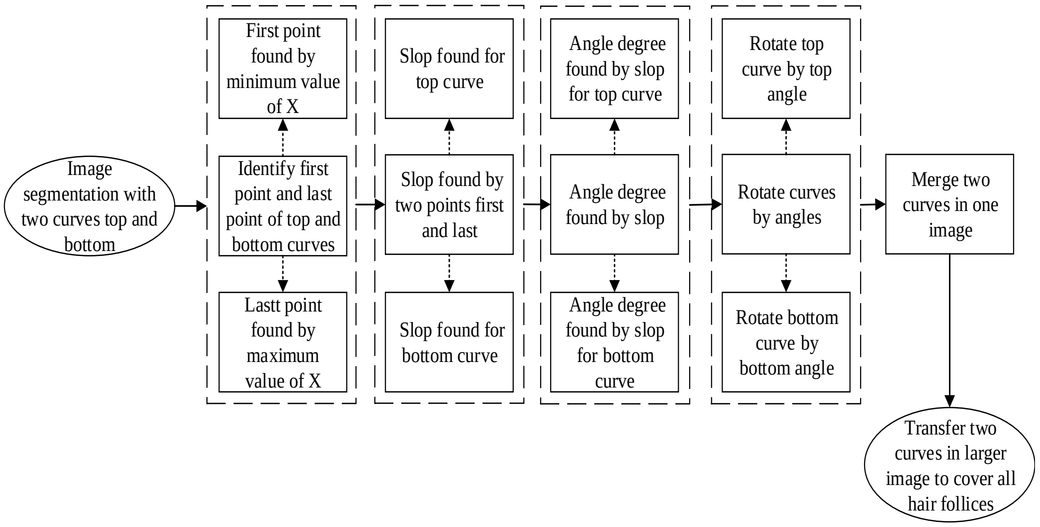


Figure 6.7 Block diagram for automatic aligning the epidermis layer

6.2.2.3 Segmenting the Hair follicle

At this stage, the hair follicle of the epidermis layer was segmented by using the proposed algorithm. After applying the alignment algorithm of the epidermis layer (top and bottom curve), which is associated with the hair follicle, the alignment image was used as the input for the next stage, which was to segment the hair follicle. The algorithm for segmenting the hair follicle begins with thickening the top curve of the epidermis layer. This step was required to investigate the thickness, which will be used to partition the hair follicle later. The following Pseudocode 6.1 is clarify the steps to segment the hair follicles.

```
Step 1:  Thicken the top curve by 100 pixels.

Step 2:  Dilate and erode the top curve by 2 pixel with disk shape.

Step 3:  Invert the output curve of step 2.

Step 4: Multiply each pixel of the top curve ( step3) with each pixels in the bottom
          curve, then the hair follicles are segmented.
```

Pseudocode 6.1 Steps for segmenting hair follicle

Conducting several experiments on the segment of the top curve to obtain an efficient division for hair follicle so that is does not affect the accuracy of segmentation of hair follicles. While we were testing the segmentation by visual effects, we found that 100 pixels is an appropriate representation for thickening the curve whilst retaining the curved shape, and to divide the hair without affecting the accuracy of the follicles, as displayed in Figure 6.8(A, B). Afterwards, the disk 2 of morphological dilation and erosion was used to close unwanted pixels inside the object in the image, as shown in

Figure 6.8B. The filtering process was used on the two images, first processing the image which has the top curve, as shown in Figure 6.8B, and then the second image which has the bottom curve that consists of the hair follicles, as displayed in Figure 6.8C. Figure 6.8D shows the inversion of the pixels of the top curve to 0 and multiply each pixel with each pixels in the bottom curve, then the hair follicles were segmented.

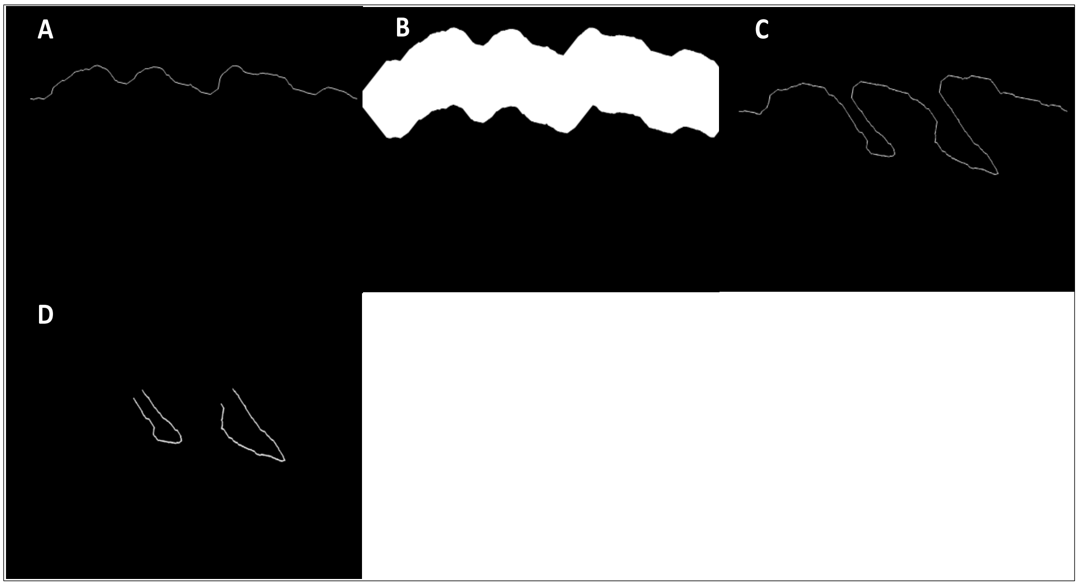


Figure 6.8 Automatic segmentation of hair follicle of H&E images. A) Top line of epidermis layer. B) Thicken the line by adding 100 pixels around it. B) Align bottom line of epidermis layer. D) Hair follicle segmentation

6.2.3 Quantification of Hair Follicle Orientation

Having segmented the hair follicle in the epidermis layer, we then implemented our orientation quantification algorithm on the hair follicle which was able to label each follicle, as shown in block diagram in (Figure 6.9) and (Figure 6.10). The block diagram in Figure 6.10, show this algorithm steps.

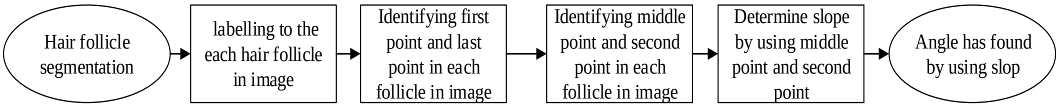


Figure 6.9 Block diagram for quantifying the orientation of hair follicle

This algorithm determines the first point and last point for each follicle (Figure 6.10), by identifying the coordinate points for the first pixel per labelled follicle, represented by the minimum x-coordinate, and similarly the follicle’s last point is represented by the minimum y-coordinate. Note that the origin of the image in MATLAB is at the top left corner. Next, the middle point of the straight line joining the first and the last point is determined. We then determine the length of the follicle as the maximum Euclidean distance between all points in the follicle. This will be estimated as the distance between the middle point and the second point which represents the tip end of the follicle. The second point is the furthest point on the follicle boundary from the line between first and last points, see Figure 6.10 as an illustration. Finally, the orientation of the hair follicle is determined by the angle that the line from the middle point to the second point makes with the x-axis which represent the alignment line of the epidermis. Accordingly, the slope of this line, represents the orientation of the hair follicle, see Figure 6.11.

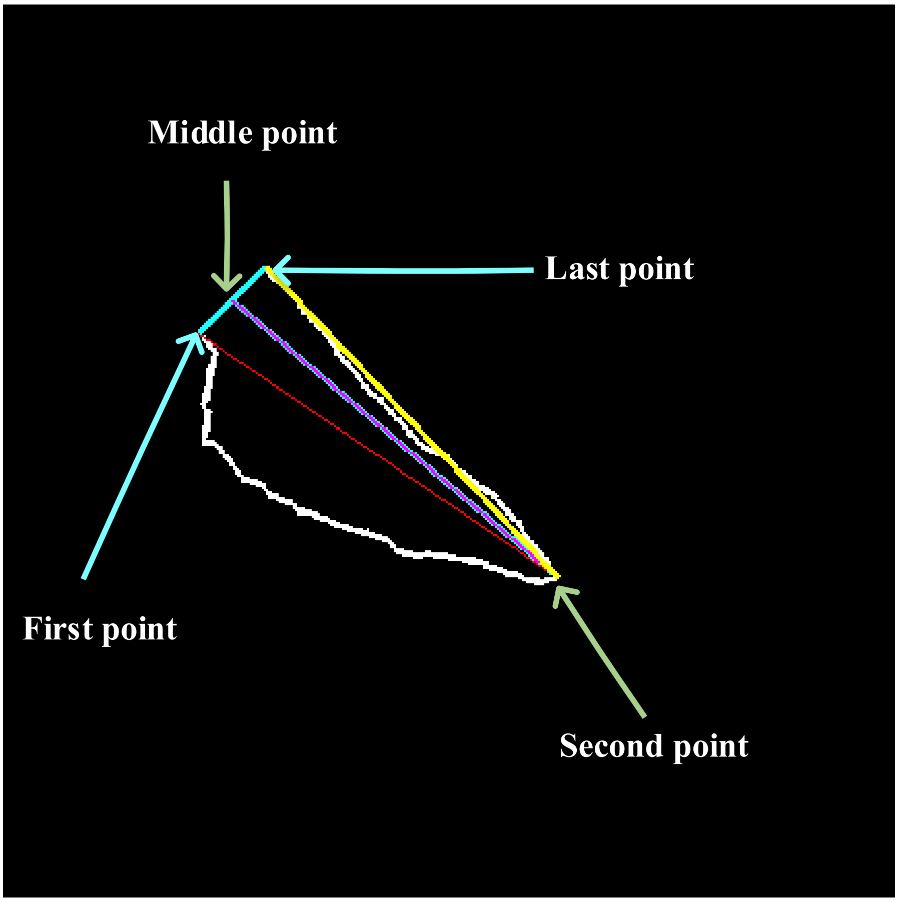


Figure 6.10 Determining the middle point and second point for each hair follicle

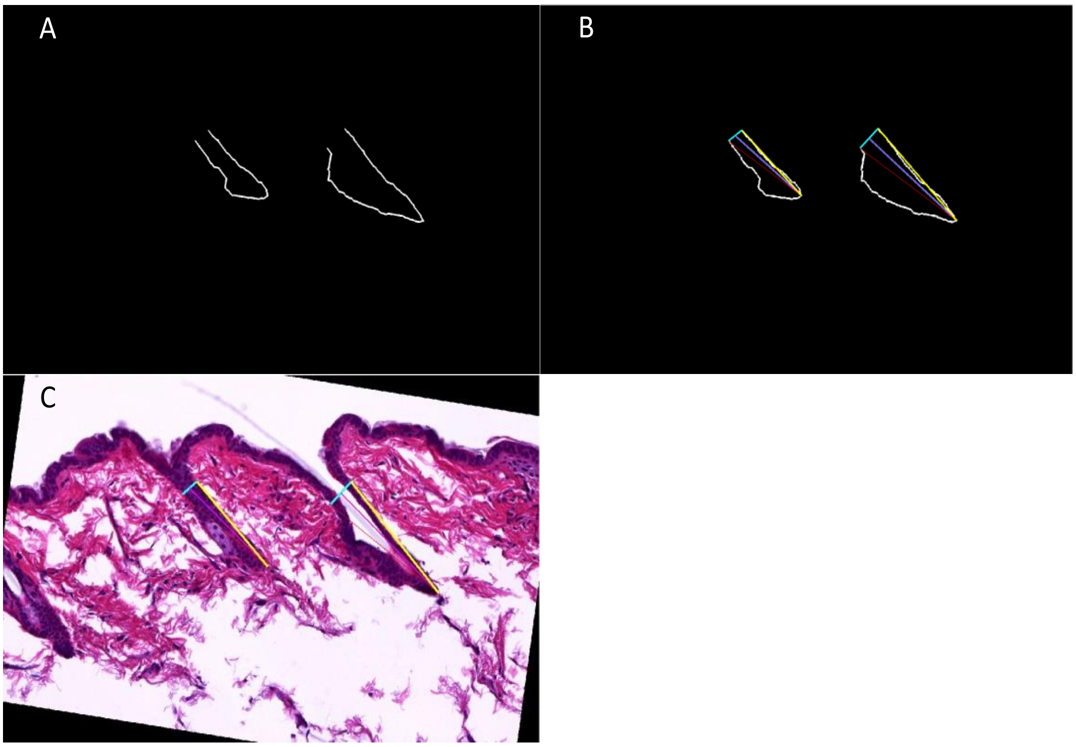


Figure 6.11 Automatic quantity the orientation of hair follicle in epidermis layer of H&E images. A) Hair follicle segmentation. B) Determined the first point and last point for each hair follicle gland. C) Quantify the orientation of hair follicle to angles by using the slop for each hair follicle gland

6.3 Experiment Results and Analysis

We conducted two sets of experiments, the results of which are presented in this section. The first experiment is designed to compare the automatic hair follicle orientation measurements to those obtained manually, while the second experiments is designed to identify genes responsible for changes in the hair follicles orientation using H&E images of WT and Mutant type mouse skin tissue. Due to the difficulty of manually counting the hair follicles in all the experimental images, in these experiments a random sample of 100 images were selected by an expert to be used for ground truth comparison. To examine consistency among the measurements, throughout these experiments, we calculated R2, ARL, MAE and standard deviation of absolute error between the ground truth and the automatic measurements (σ) (Section 3.4.2).

6.3.1 Manual vs. Automatic Measurements

Two comparisons of manual Vs. automatic hair follicle measurements were conducted. In the first one, we compare the accuracy of segmenting follicle from epidermis layer, while the second one deals with the angle of follicle orientation measurements.

6.3.1.1 Manual vs. Automatic Method of Hair Follicle Segmentation

To assess the accuracy of the hair follicle segmentation, between the automatic segmentation and the ground truth, we counted the number of segmented hair follicles in the test images for both automatic and manual segmentation. The mean (μ) of the number of hair follicle segmentation and the standard deviation (σ) across all testing images were calculated. The results are displayed in Table 6.1. From the table above it is clear that our automatic method is highly successful in segment hair follicles from epidermis layer. In Figure 6.12, there are two lines of bars, the red bars representing manual count of the follicles, and the blue bars are the automatic count. It is clear that two bars were close in parallel during 100 images that confirming good success. To examine consistency among the measurements, we calculated R^2 (see Section 3.4.2), and it was 80%, shows the close correlation between the manual and the automatic in most cases.

Table 6.1 Comparing mean and standard deviation of follicle counts automatic Vs. ground truth

Measure	Automatic		Manual	
	μ	σ	μ	σ
Hair follicle segmentation	1.92	0.91	2.18	1.05

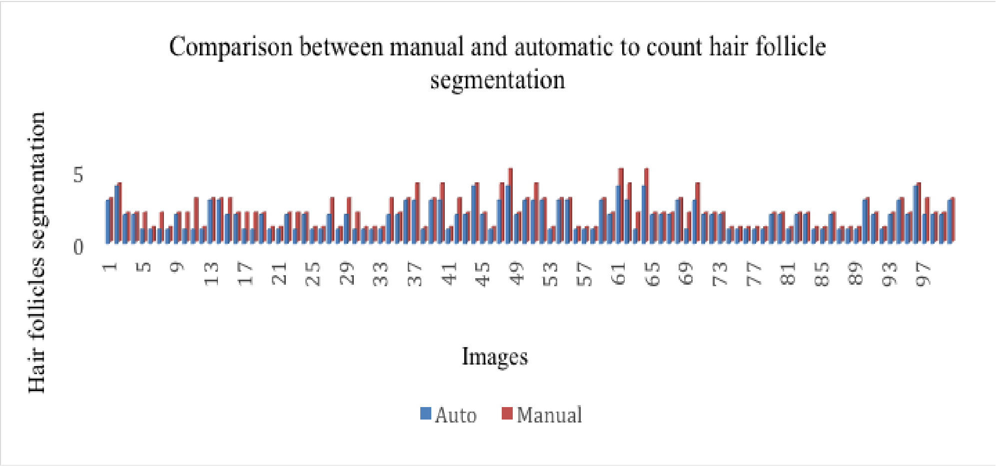


Figure 6.12 Manual and automatic comparison to hair follicle segmentation

6.3.1.2 Manual vs. Automatic Method of Hair Follicle Orientation

The proposed algorithm quantifies the orientation of the hair follicles in term of the angle it make with the angles. Each of the experimental images consists of more than one hair follicle, some images have two or three hair follicles. Therefore, in our algorithm, we use median value of the computed angles for all follicles in the input image. In this case, the median angle sets the orientation in the image as that of the middle value. For example, when we quantify three hair follicle orientations in an image by three angles, if an error occur in one angle value while the other two angles were correct, then the middle value of the orientation of hair follicle is selected based on median (see Figure 6.13 (A, C)).

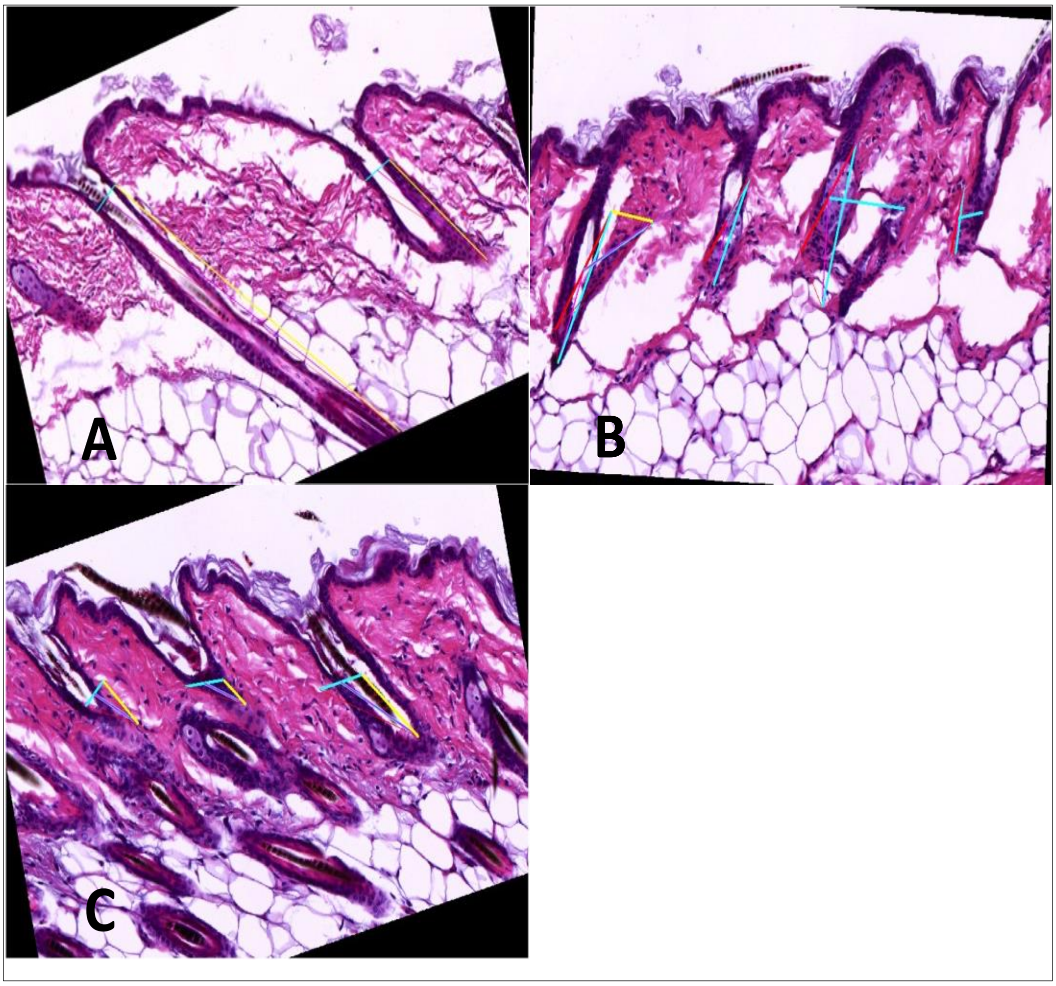


Figure 6.13 Successful and unsuccessful follicle segmentation. A) Good segmentation. B) Bad segmentation. C) Dual segmentation, the middle follicle wasn't accurate but the other were accurate

Furthermore, our experiments have shown that our algorithm results in choosing the accurate orientation value with a high probability as indicated by the values of R^2 (see Figure 6.14), ARL, MAE and σ, see Table 6.2.

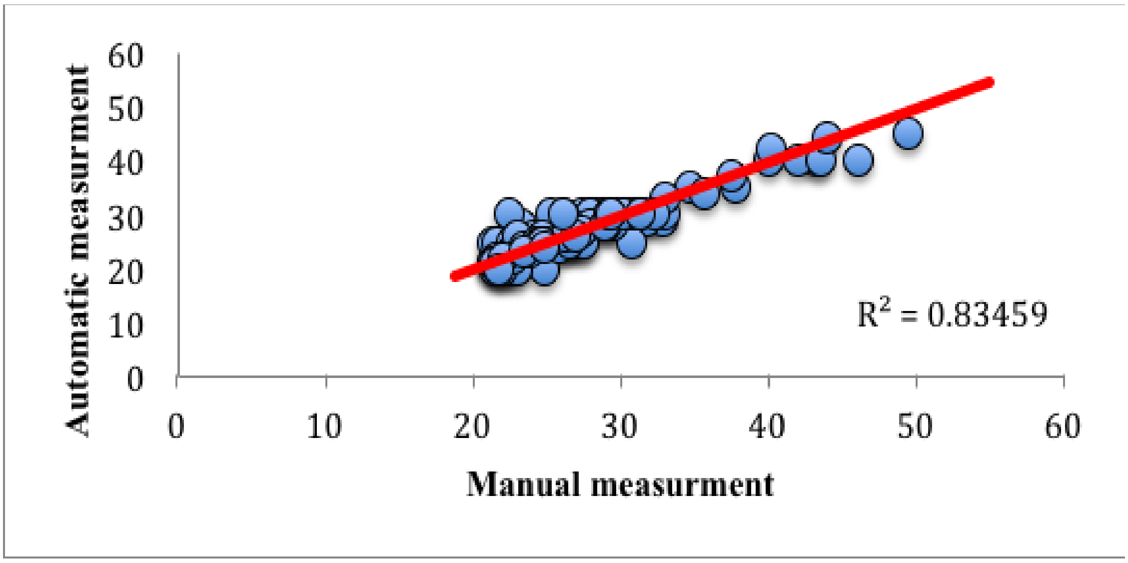


Figure 6.14 Correlation of manual versus angles of hair follicles measurements

Table 6.2 Manual versus automatic measure angles for hair follicles orientation

Measure	R^2	ARL	MAE	σ
Hair follicle orientation	83%	45.8°	0.018	0.18

However, after assessing the correlation between the manual and automatic measurements, we need to examine the closeness between them using the BA analysis (Section 3.4.2).

Table 6.3 shows the lower and upper LoA for all measurements.

Table 6.3 Manual versus automatic measures the angles of the hair follicles

Measure	Upper LoA	Lower LoA
Hair follicles orientation	4.98	-5.24

Figure 6.15 shows the BA analysis scatter plot for measuring hair follicles orientation, which is display closeness between automatic method and ground truth (see section 3.4.2 for further details) by using 100 images randomly. Again, the results in Table 6.3 confirm the conclusion of Table 6.2 and show that the hair follicle orientation was quantifying successfully, the reason behind that is LoA range for angles of the hair follicle orientation in the epidermis layer, which is start from -5.25 to 4.24 and most of the points located inside the LoA range, which indicated to the correct cases as shown in Figure 6.15. While some points located outside LoA range (Figure 6.15), which indicated as incorrect cases.

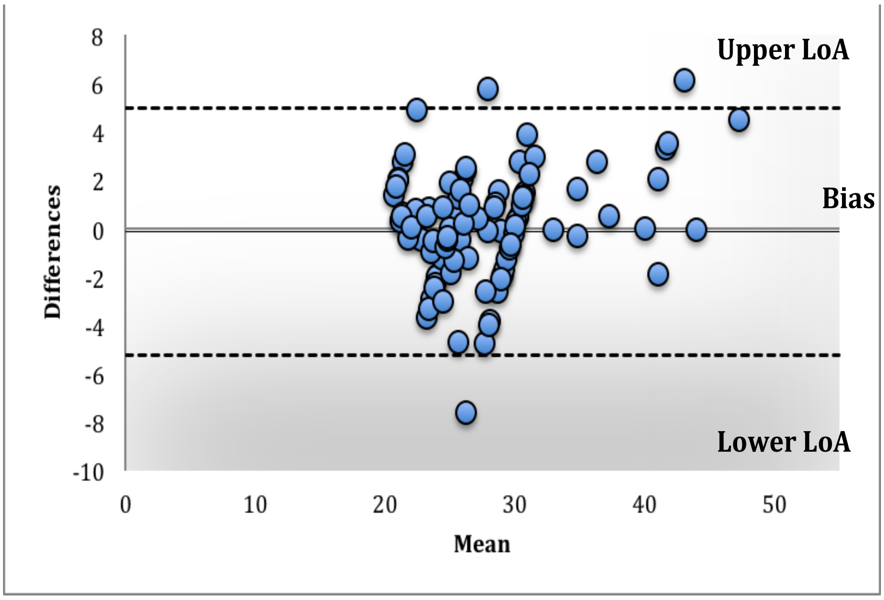


Figure 6.15 Bland Altman analysis for manual versus automatic for measuring hair follicles orientation

6.3.2 Genes Identification

In this section, we report the results of applying the hair follicle counting and orientation methods on the whole experiment dataset of 4000 H&E images. The goal is to identify the genetic basis of abnormalities in the orientation of the hair follicles i.e. what are the genes that may lead to increase or decrease in the orientation of angles? To do this, first needs to establish what is normal and what is not when it comes to hair follicles orientation in mice skin tissue. This is typically done based on the RR approach to identify the LB and UB of the nuclei counting obtained from WT animals (see Section 3.4.1 for further details). Subsequently, the process of identifying genes associated with abnormalities in orientation in the epidermis layer goes through the following two steps:

Step 1

Hair follicles orientation of every mutant mouse were compared against the relevant LB and UB to establish whether the measurement falls within or outside the RR for WT images. The LB and UB of the RR, obtained from the orientation of follicles of WT (normal) cases, were shown to be (-25 to 34).

Out of 4000 mutant mice images analysed, 3200 (80%) were found to be within the RR established for the WT measurements, and 800 mice (20%) of mutants fell outside the RR interval. However, it is important to highlight that not all samples that fall outside the RR are necessarily true hits. Arguably the above number includes many false cases due to two main reasons. Firstly, any incorrect follicles detection, may lead to incorrect measurement of angles. Secondly, even when the follicles of an image are correctly measured as abnormal, the total number of abnormal orientation angle cases for the associated specific gene remains less than the hit rate threshold (see Section 3.4.1 for further details). Hence, our analysis requires an additional experiment step.

Step 2

To filter out the false cases, we calculated the so-called hit ratio of each mutant line Out of the 800 outlier cases, only 80 cases were considered as true hits (i.e. hit ratio > 60%) as shown in Table 6.4 where the "-" mark is used to indicate that the ratio is below 60%, we did same steps of Chapter 5 Section 5 to identify interesting genes.

Table 6.4 Possible gene hits. In hair follicles in the epidermis layer

Name of allele	Orientation of the hair follicles	
	Phenotype	Hit Ratio
1300010M03Rik/1300010M03Rik	>UB	86%
Traf3ip3/Traf3ip3	>UB	71%
Aldh18a1/+	>UB	100%
Adam3/Adam3	>UB	86%
Tmem98/+	Normal	-
Lrrc67/Lrrc67	Normal	-
Secisbp2(c)/Secisbp2(c)	>UB	71%
Actr6/+	>UB	71%
Kcnv2/Kcnv2	>UB	-
Inpp1(b)/Inpp1(b)	>UB	100%
Lix1l/+	<LB	86%

The table above shows a list of 7 genes that could be potentially responsible for abnormalities in the orientation of hair follicles in the epidermis layer, 6 genes show increased orientation and 1 gene show decreased orientation in the epidermis. To some extent, this confirms the effectiveness of our developed method in identifying the list of genes responsible for changes in the orientation in the epidermis layer. Figure. 6.16 shows an example of these interesting genotypes, particularly Traf3ip3/Traf3ip3, which yields increased orientation angles while Lix1l/+ is associated with a decreased orientation angles of hair follicles.

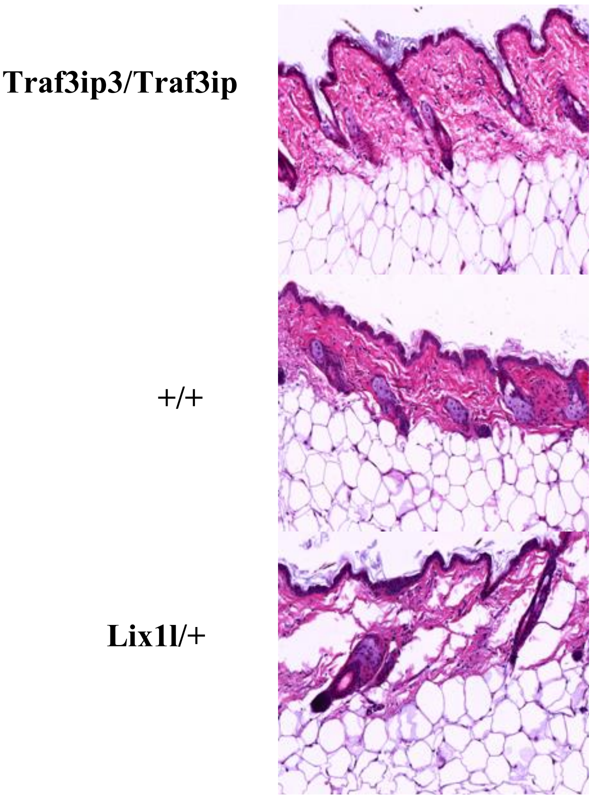


Figure 6.16 Images of genotypes with that influence of changes in orientation of hair follicle compared to wild type from development data set (4000 images at 20X) magnification)

Experimental results on a large set of H&E images of 500 WT and 4000 mutant mice have demonstrated the effectiveness of the proposed solution in separating images of abnormal orientation of the hair follicles from those of normal ones.

Analysing the results led to identifying 7 genes associated with abnormality in orientation of hair follicle, whereby (Aldh18a1, 1300010M03Rik, Adam3, Secisbp2(c), Actr6 and Inpp1(b)) are related with increased orientation and only one gene (Lix1l) related with decreased orientation. These results were positively confirmed by domain experts.

6.4 Discussion and Conclusion

Hair follicles play an important role in the histological changes in mouse skin tissues that can be observed in their microscopic images. The disorder in the follicle is caused by several cases, including mutation phenotype and health conditions such as inflammation of the follicle. In this chapter, we investigated and developed automatic methods to segment and quantify the orientation of hair follicles in epidermis layer. Testing the reliability and effectiveness of the developed schemes with a sufficiently large dataset of H&E microscopic skin images of WT and Mutant type mice, has demonstrated a remarkable success.

Having identified the technical challenges associated with the task computing hair follicles orientation, our systematic approach helped dealing with these challenges and led to several contributions:

1. Highly successful segmentation of the outer sublayer of the epidermis layer that include the hair follicle glands from which the follicle glands are easy to isolate.
2. Defined a simple procedure to align the segmented sublayer of the epidemies in normalise manner that could be applied to any image and facilitate an effective uniform definition of follicle orientation.
3. Designed a procedure for evaluate the orientation of each segmented follicle and output a single orientation value representing all follicles in skin section image.

Our experimental work with a random sample of 100 images, demonstrated an excellent correlation between our automatic schemes and the ground truth determined manually by domain expert. The success and effectiveness of the developed algorithm was further manifested by the experiments that were designed to identify the genetic phenotype of skin abnormalities where 7 genes were identified, in terms of the changes of follicle orientation, and confirmed by domain experts.

Chapter 7 : Automatic Quantification of Epidermis Curvature in H&E Stained Microscopic Skin Image of Mice

The skin can reveal evidence of inflammation, hyperplasia, connective tissue disorders and underlying metabolic changes resulting from local and systemic influences. In the last two chapters, we investigated and developed effective algorithms for nuclei segmentation and counting and determining hair follicle orientation, and demonstrated the effectiveness for identifying genes that may cause some of the skin abnormalities. In this chapter, we conclude our investigations by developing automatic quantification of changes to another important skin structure parameter for the same end points of the last two chapters. Changes in the epidermis layer *curvature* is known to be associated with many skin disorders, such as ichthyoses and generic effects of ageing (S.C., 2002). Hence, the focused contributions of this chapter are:

- Quantify the curvature in the epidermis layer, and also
- Assess changes in the curvature to identify interesting gene.

Naturally, our methods starts by using the epidermis layer segmentation algorithm developed in chapter 4, and the method used in last chapter to segment and align the two bounding curves of the outer epidermis sub-layer. The epidermis curvature-based parameter will be quantified the amount of deviation of the bottom curve length from the length of alignment line, using the sum of distances between the curve’s points and the alignment line.

The rest of this chapter is organised as follows: In Section 7.1, the problem statement of this chapter is described. Section 7.2 presents our proposed automatic method. Section 7.3 reports and analyses the results of experimental work to identify the genes associated with curvature changes in the epidermis layer in mice skin images. Section 7.4 discusses a source of errors in the proposed solution with analysis of the genetic experimental results. Finally, conclusions from these investigations are summarized in Section 7.5.

7.1 Problem Statement

Several studies showed that the curvature of rete ridges can be varied in ageing or obesity (Honda *et al.*, 2007). Other studies found that several skin disorders are associated with changes in epidermal junction characterization, such as psoriasis (Kurugol *et al.*, 2011). Quantifying the disorder in the epidermis layer refer to the accurate changes in curvature in this layer from an H&E image of a mouse. Automatic curvature quantification is concerned with effective computer based solutions to accurately measure the change in the epidermis curve. Manual measuring of the epidermis curvature is time consuming, prone to errors, and a tedious process. This is because the outer boundary of the epidermis is quite complex in that it is non-smooth even at tiny scales. Besides the high throughput nature of the task due to the huge number of H&E images that are regularly generated from the problem domain (i.e. MGP). Furthermore, staining often causes difficulties in exact epidermis location, resulting in error segmentation of the region of interest by inexperienced observer. Figure 7.1 shows an example of staining issue in epidermis layer which displays the changes in locations of epidermis. In our proposed method to overcome this issue by introduced automatic method to give a deterministic and consistent result.

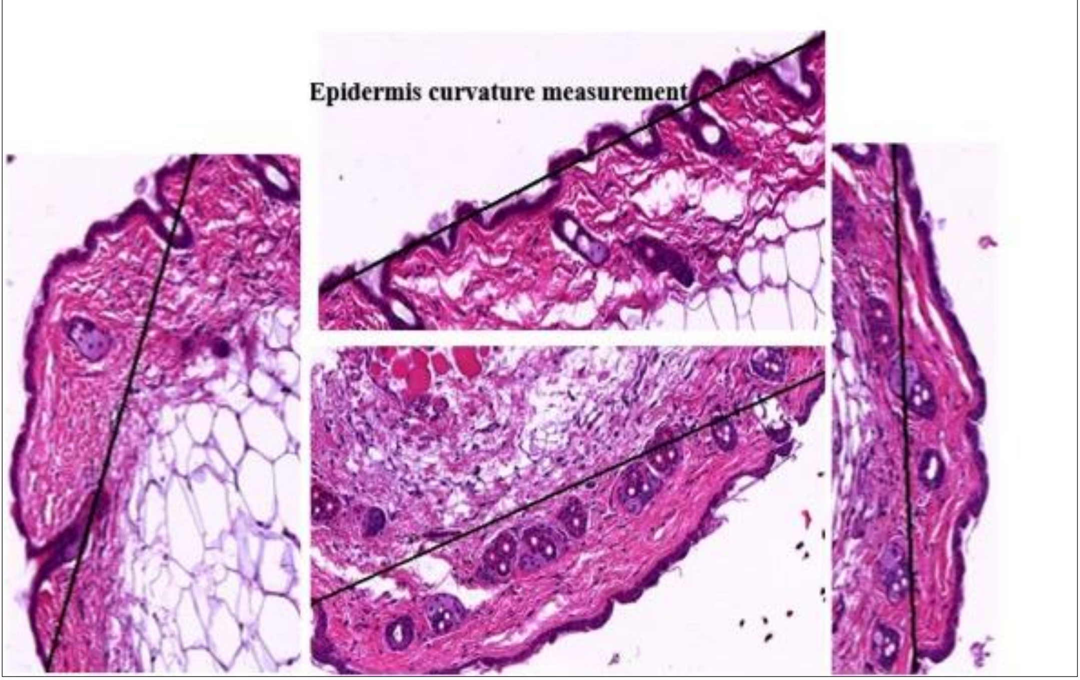


Figure 7.1 Manual measuring of epidermis curvature to different images

Automatic quantification of epidermis curvature in mouse images faces the following challenges:

1. The stained tissues in many images of our dataset suffer from loss of parts of certain layers such as epidermis layer, which cause overlap with the neighbouring layer (i.e. the dermis) as shown in Figure 7.2.

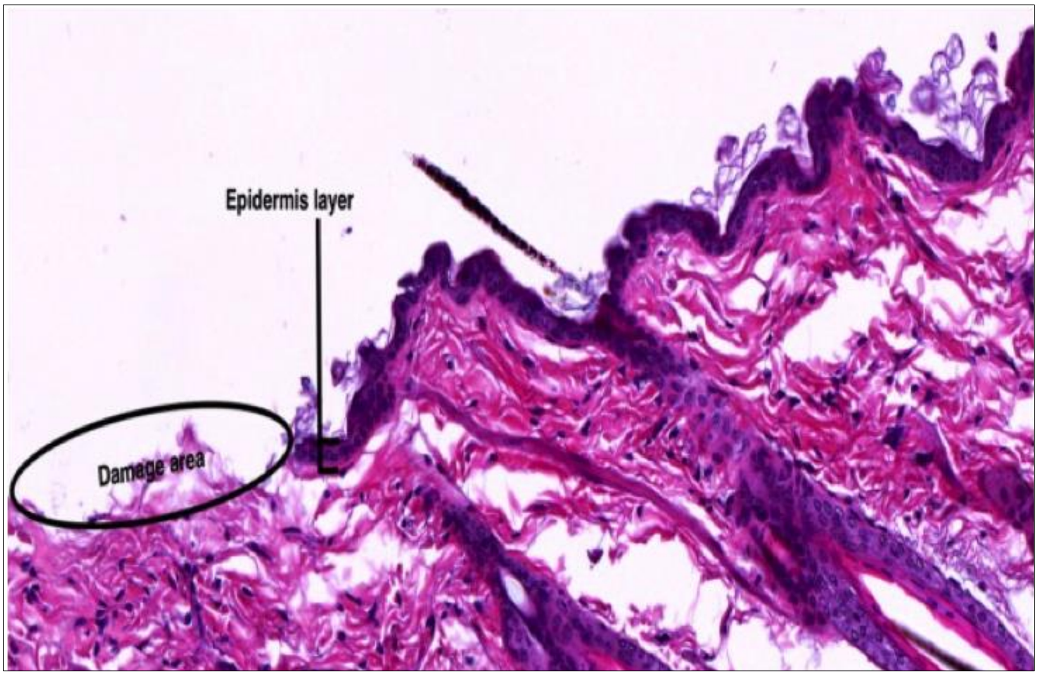


Figure 7.2 Damage in epidermis layer

2- The epidermis layer consists of two sub-layers, the cornified and basal (more details in Chapter 4). The relatively low resolution of image dataset (of 20X magnification) causes difficulties in distinguishing the Cornified sub-layer, which leads to overlap between the cornified and the epidermis, i.e. subtle differences between intensity values of nearby pixels. In Figure 7.3, displays the overlap between cornified and epidermis.

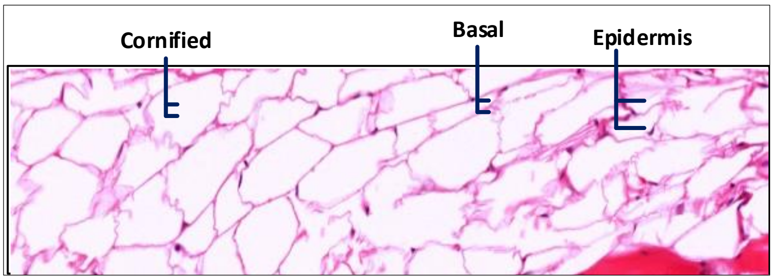


Figure 7.3 H&E image with overlap between basal and cornified layers. Epidermis layer touching the edge image

3- The proximity of the curvy shape of the epidermis layer in most our dataset images, being too close to the top edge of the images, confuses the segmentation process. This problem is illustrated too in Figure 7.3.

7.2 The proposed Method

In this section we shall describe our proposed method for automatically computing the curvature along the epidermis layer as a significant factor for identifying genes associated with changes in curvature values. Before presenting the algorithm, we need to point out that the term “curvature” used here is loosely related to the mathematical concept of curvature determined with the second derivative of the function that fits the curved shape that modelling the points of the epidermis. Unfortunately, determining the best curve fit of the epidermis is a very challenging task due to the very large number of such points besides many other technical challenges relating to variation in alignments of different sections and influence of image magnification. Hence, we shall quantify the epidermis curvature in terms of the way it winds and curves around a straight line that tightly connects the epidermis (i.e. around its skeleton). This approach was developed to reflect the way life scientists measure this parameter.

The proposed algorithm works in three key stages described below as shown in Pseudocode 7.1, the first two steps describe the computation of the curvature while the last step is concerned with quantifying the change in epidermis curvature between two samples. For the sake of clarity we illustrate the algorithm by the example in Figure 7.4.

```
1. Segment the epidermis layer by the procedure described in Chapter 4, and segment its outer band bounded by the upper and lower curves described in chapter 6.
2. Quantify the curvature of the epidermis layer by measuring the difference between the epidermis outer band and a straight line (an artificial reference line) connecting the two far sides of the epidermis outer band.
3. Use the output curvature measurements extracted from the images in the large dataset of H&H skin images of mutant mice to identify a list of genes responsible for changes in the epidermis curvature. (The range of computed curvature over all images is expected to distinguish the curves in the wild type from the mutant types)
```

Pseudocode 7.2 Steps for epidermis curvature quantification

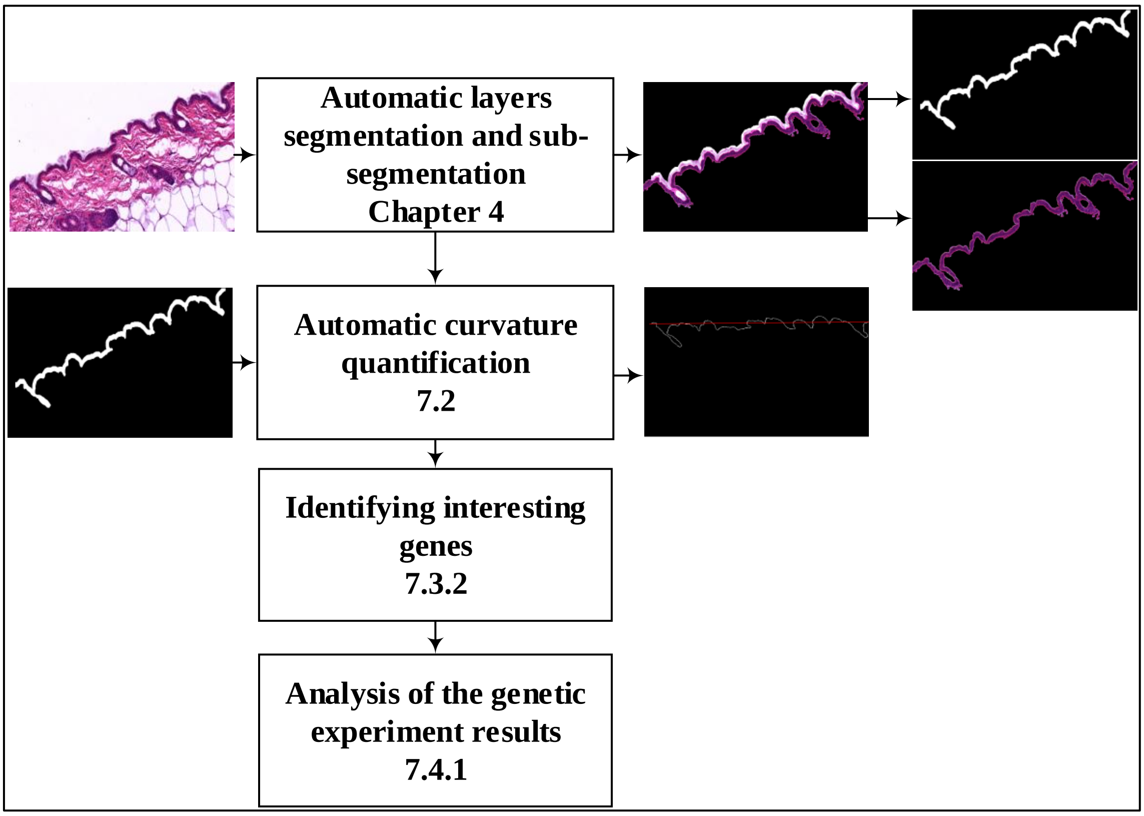


Figure 7.4 Epidermis Curvature quantification algorithm for identification of genes causing significant changes

We shall now, give details of the way the 2nd step of our algorithm to quantify the curvature of a segmented epidermis by measuring the difference between the epidermis curve and a straight line (an artificial reference line) connecting the two far sides of the epidermis edge.

<u>Curvature computing Algorithm</u>

Input -the border band of the epidermis.

Step1. Select the bottom border of the epidermis mask (cornified sub-layer) - Figure (7.5C)

Step2. Connect the far two sides using a theoretical reference line (the red line in Figure (7.5D), and determine its equation in the form:

$$Ax + By + C = 0 \qquad \textbf{\textit{(eq7.1)}}$$

Step3. Starting from the initial point of the selected border curve, calculate the array of distances between the points (u,v) on the curve and the reference straight line using the formula:

$$D(u,v) = \frac{|Au+Bv+C|}{\sqrt{A^2+B^2}} \qquad \textbf{\textit{(eq7.2)}}$$

For details see (Libby, 2017).

To reduce the effect of noise, if D(u,v) is less than a chosen threshold then set it to 0.01

Step4. Calculate and output the mean of the D array as the epidermis curvature.

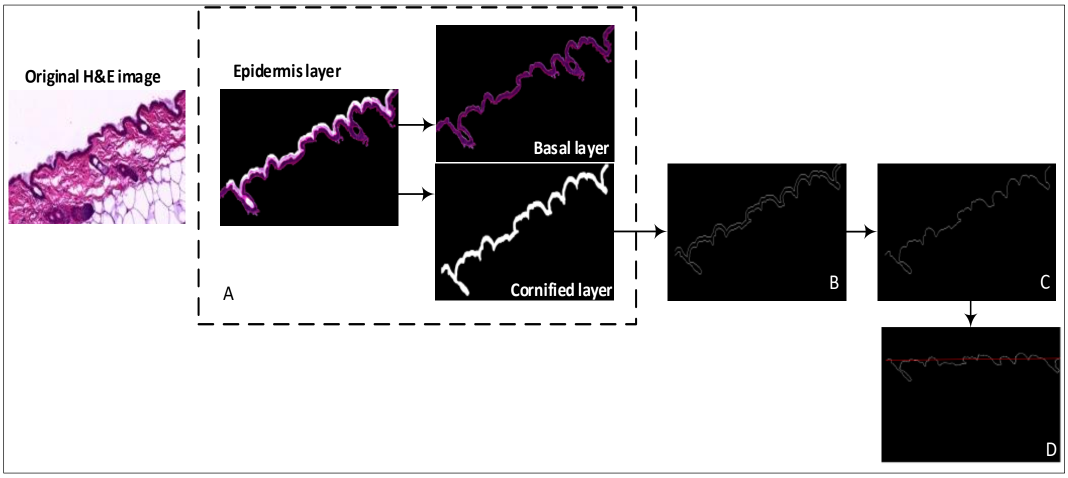


Figure 7.5 Representative images for each step of the automated image analysis method to find the curvature in the epidermis layer. A) Epidermis layer segmentation and sub-segmentation (cornified layer). B) Border for the cornified layer. C) Bottom line segmentation. D) Theoretical red line on bottom curvature

7.3 Experiment Results and Analysis

Two sets of experimental results are presented in this section. The first set is aimed to compare the automatic measurements with manual. While the second set of experiments is devoted to the identification of genes responsible for changes in curvature.

7.3.1 Manual vs. Automatic curvature quantifications

A random samples of 100 images has been selected by domain expert to be used as a ground truth due to the difficulty of measuring the curvature manually for the entire large dataset of H&H mouse skin images. To assess the accuracy of measuring the epidermis curvature, the distance between the automatically measured and the ground truth points was calculated in micrometer (mm). The scale for the conversion from pixel to micrometer for the mouse 20X image adopted in this research is 0.5 mm for every pixel (Mann *et al.*, 2005).

As explained above, we quantify the curvature of the epidermis layer by measuring the distance between each points and reference line, which is connected between the first and the last point in the epidermis edge mask. In the manual measuring, the distances between all peaks and the reference line in the epidermis layer. In both cases, the averages of the corresponding distances are used as the curvature measure.

To evaluate the proposed curvature method of the epidermis layer versus the manual measurements taken by a domain expert. In the evaluation, we check the correlation between the manual and automatic measurements using R^2 and ARL (see Chapter 3 for further details). Figures 7.6 shows the scatter plot for the manual versus automatic measurements for curvature based on distance in the epidermis. The values and figure of ARL and R^2 show the close correlation between the manual and the automatic in most cases.

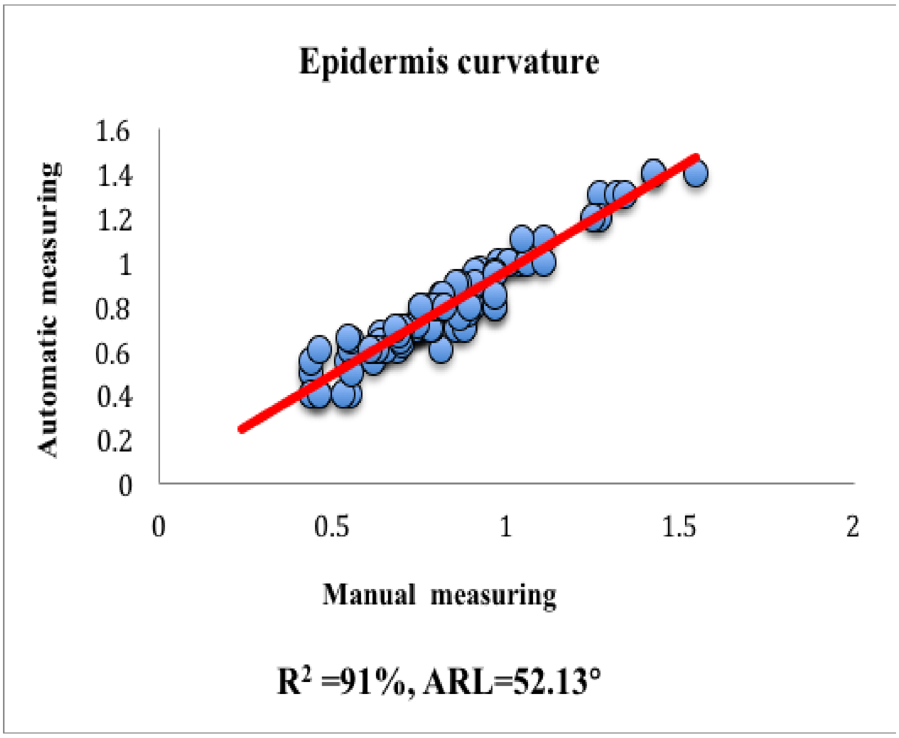


Figure 7.6 Correlation of manual versus automatic measurement

Furthermore, Table 7.1 reports the comparison results based on two parameters, namely R^2, ARL between the ground truth and the automatic measurements (see Chapter 3 for further details). Again these results show that the proposed method is successful to quantify the curvature in epidermis layer.

Table 7.1 Manual versus automatic curvature measuring in the epidermis layers

Measure	R^2	ARL
Curvature measuring in the epidermis	0.91	52.13°

However, after assessing the correlation between the manual and automatic measurements, we need to examine the closeness between them using the BA analysis (see Chapter 3). Table 7.2 shows the lower and upper LoA for all proposed measurements.

Table 7.2 Manual versus automatic for curvature measuring epidermis layers

Measure	Upper LoA	Lower LoA
Curvature measuring in the epidermis	0.087	-0.061

Figure 7.7 shows the BA analysis scatter plot for measuring the curvature in epidermis layer, which is display closeness between automatic method and ground truth (see section 3.4.2 for further details) by using 100 images randomly.

Again, the results in Table 7.2 confirm the conclusion of Table 7.1 and show that the curvature of epidermis layer was quantifying successfully, the reason behind that is LoA range for the mean of the distances of the epidermis layer, where the range is (-0.061 to 0.087) and most of the points located inside the LoA range, which indicated to the correct cases as shown in Figure 7.7. While some points located outside LoA range (Figure 7.7), which indicated as incorrect cases.

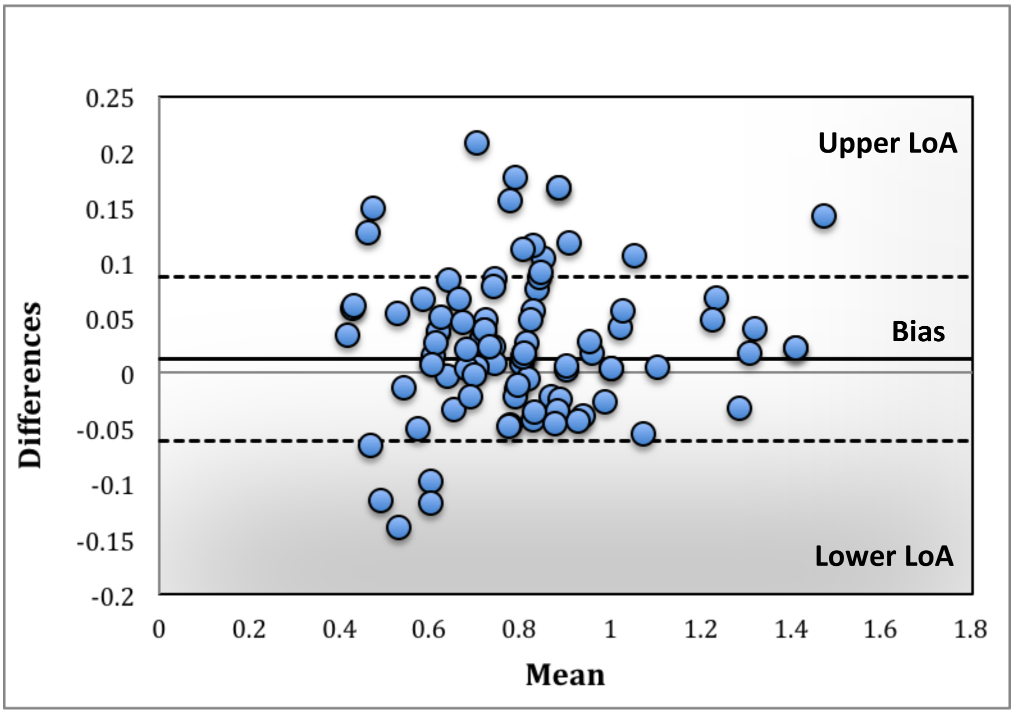


Figure 7.7 Bland Altman analysis for manual versus automatic for curvature measuring in the epidermis layer

7.3.2 Genes Identification

In this section, we report the results of applying the proposed method on the whole dataset of 5700 H&E images. The ultimate goal of the experiment is to identify the genetic basis of abnormalities in the epidermis curvature i.e. what genes may cause an increase or decrease in the value of curvature of the epidermis? We used the RR approach to identify the LB and UB of the curvature measurements obtained from WT animals (see Section 3.4.1 for further details).

After that, the process of identifying the genes associated with abnormalities in epidermis curvature follows the two stages below.

Stage 1

Curvatures of every mutant mouse were compared against the relevant LB and UB to establish whether the measurement falls within or outside the RR. Similar to WT images. The LB and UB of the RR (0.60 to 1.26), which obtained from the curvature of WT (normal) cases. However, it is important to highlight that not all samples that fall outside the RR are necessarily true hits. Arguably the above number includes many false cases due to two main reasons. First is related to any incorrect layer segmentation, which leads to incorrect measurement of curvature. Second reason is that although the case is correctly measured as abnormal but this case might belong to aspecific gene where the total number of cases of abnormal curvature for it is less than the hit rate threshold (see Chapter 3 for further details).

Stage 2

To filter out the false cases, we calculated the so-called hit ratio of each mutant line (see Chapter 3 for details). Out of the 780 outlier cases, only 90 cases were considered as true hits (i.e. hit ratio > 60%) as shown in Table 7.3. Here the “-” mark is used to indicate that the ratio < 60%.

Table 7.3 Possible interesting genes

Name of allele	Curvature in the epidermis layer	
	Phenotype	Hit Ratio
Abcd1/Abcd1	>UB	75%
Slc44a5/Slc44a5	>UB	100%
Mgst3(b)/Mgst3(b)	>UB	100 %
Tmem98/+	>UB	100%
1190002H23Rik/1190002H23Rik	>UB	100%
Plekhg1/Plekhg1	>UB	100 %
Amotl1/Amotl1	>UB	95%
1500011B03Rik/1500011B03Rik	Normal	-
Gpc6/Gpc6	>UB	100%
Pld5/Pld5	Normal	-
Clk1/Clk1	>UB	95%
Fam96a/Fam96a	Normal	-
Usp4/Usp4	Normal	-
Tm9sf4/Tm9sf4	>UB	100%
Coq9/Coq9	<LB	80%
Pik3cb/+	<LB	90%
Pfn1/+	<LB	77%
2010107G12Rik/2010107G12Rik	Normal	-
Psat1/+	Normal	-
Blzf1/Blzf1	Normal	-
Secisbp2/+	<LB	92%
Orc3/+	<LB	89%
Pfn1/+	<LB	87%
Mapk10/Mapk10	<LB	81%
Anp32e/Anp32e	<LB	78%
Farsa/+	Normal	-

The table above shows a list of 18 genes that could be potentially responsible for abnormalities in the curvature in the epidermis layer, 10 genes show increased (>**UB**) curvatures and 8 genes show decreased (<**LB**) curvature in the epidermis. To some extent, this confirms the effectiveness of our developed method in identifying the list of genes responsible for changes in the epidermis curvature. Figure 7.8 shows an example of these interesting genotypes, particularly Slc44a5/Slc44a5, which results in increased curvature and Secisbp2/+ that is associated with a decreased curvature.

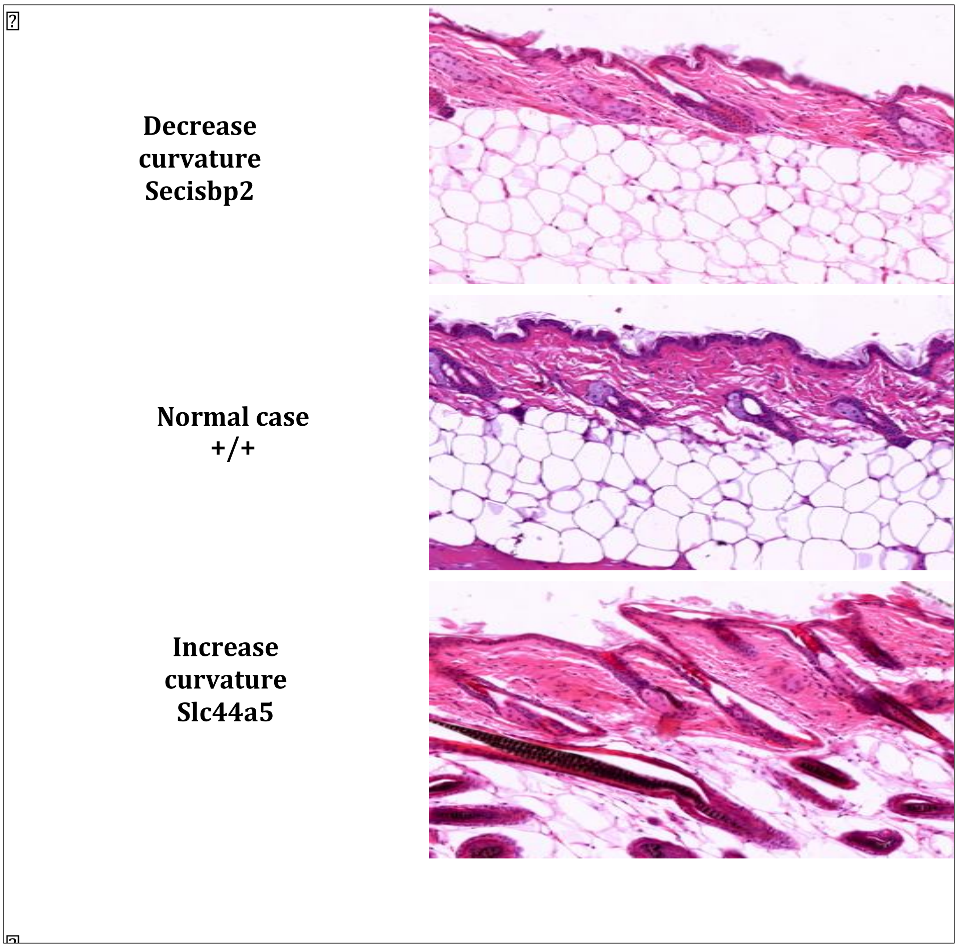


Figure 7.8 Automatic estimation of interesting genes for curvature in epidermis layer

7.4 Discussion

We developed an automated epidermis “curvature” quantification algorithm that exploits the results of segmentation of the epidermis layer and its border (edge) component, developed in the previous chapters, and computes the epidermis curvature as the mean of distances between the points of the lower bounding curve of the edge and a reference line joining the two ends of the epidermis edge. This algorithm avoids the complex mathematical computation of the curvature of a curve along its entire length while reflecting the way life scientists determine the curvature manually. Experimental work to compare the computed curvatures for a sample of 100 H&E images of mouse skin, and the ground truth as determined manually by domain expert, have demonstrated the validity of our interpretation of the curvature concept and as an effective replacement for the error-prone tedious manual approach.

This method was then applied to H&E images from the WTSI MGP to enable investigation of the effect of different genes or environmental factors such as diet, on the morphology of the epidermis, in a high-throughput screening experiments.

As it is not feasible to manually check the segmentation accuracy of 5,714 images, we took random samples of 500 images. By close manual examination, we found that 16 images out of the 500 images were not segmented successfully, as shown in Figure 7.9 the segmentation algorithm therefore achieved an accuracy about 96.8%. Due to the overlap between layers and sub-layers, such as epidermis and cornified, there was incorrect segmentation in the cornified sub-layer as illustrate in Figure 7.9.

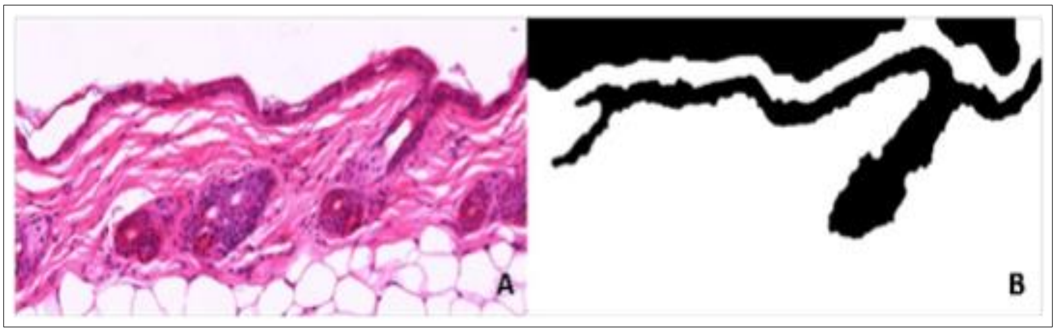


Figure 7.9 Segmentation accuracy. A) H&E image. B) Incorrect segmentation

The border segmentation of sub-layer was another challenge need to be solved, as we mentioned in Section 7.2, the ROI is touching or too close from the top edge of image. This causes split the top curve of cornified layer into many curves which is incorrect segmentation (Figure 7.10), to address that, we select the bottom segmented curve to be quantifying based on the distance and reference line.

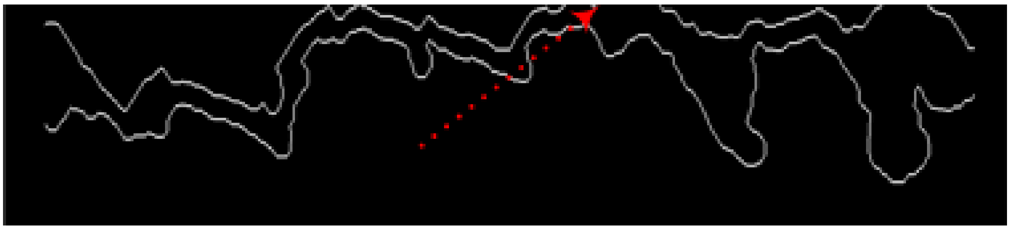


Figure 7.10 Incorrect border segmentation of cornified layer (sub-layer of epidermis)

7.4.1 Analysis of the Genetic Experimental Results

Having conducted all the genetic investigations to determine the interesting genes that cause changes to the 3 skin layers feature parameter, in this section we shall compare the results of the various experiments.

Recall that in each of the experiments conducted for the 3 parameters, we used RR to decide presence/absence of abnormalities associated with each gene. This was determined by identifying the confidence interval, where outside the interval indicate presence of abnormality (White et al., 2013).

In order to make the comparison easy, we shall display all our data distributions and charts obtained for the various experiments, i.e. analysis nuclei numbers in epidermis and dermis layer, hair follicles orientation and changes in epidermis curvature measurements.

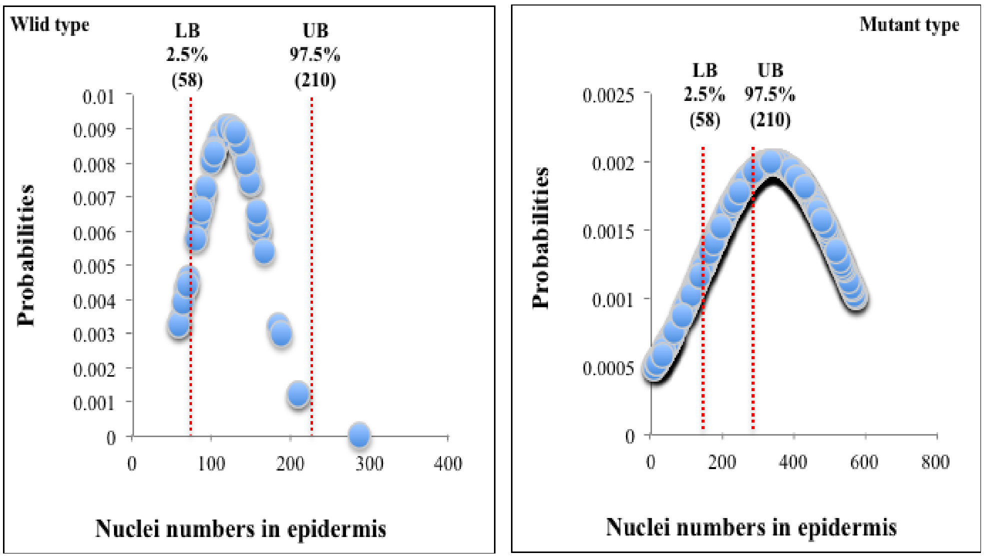


Figure 7.11 Normal distribution of nuclei in epidermis layer for wild and mutant populations with RR bound, wild type on the left side and mutant on the right side

Figure 7.11 was showed the data distribution for the nuclei counting in the epidermis layer into two types, wild (WT) and mutant. Lower Bound (LB) and the Upper Bound (UB) of the RR are determined based on wild type (58 to 210), which are pointed as two vertical red lines in Figure 7.11. Then the interesting genes has determined based on LB and UB in the mutant type as shown in the right side in Figure 7.11 (see Chapter 5 Section 5 for more details).

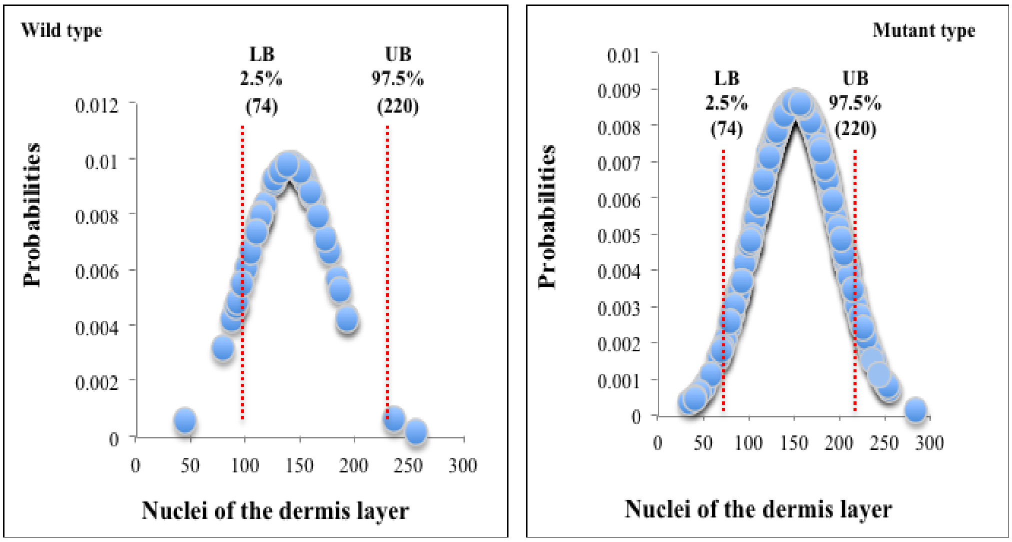


Figure 7.12 Normal distribution of nuclei in dermis layer for wild and mutant populations with RR bound, wild type on the left side and mutant on the right side

Figure 7.12 was showed the data distribution for the nuclei counting in the dermis layer into two types, wild (WT) and mutant. Lower Bound (LB) and the Upper Bound

(UB) of the RR are determined based on wild type (74 to 220), which are pointed as two vertical red lines in Figure 7.12. Then the interesting genes has determined based on LB and UB in the mutant type as shown in the right side in Figure 7.12 (see Chapter 5 Section 5 for more details).

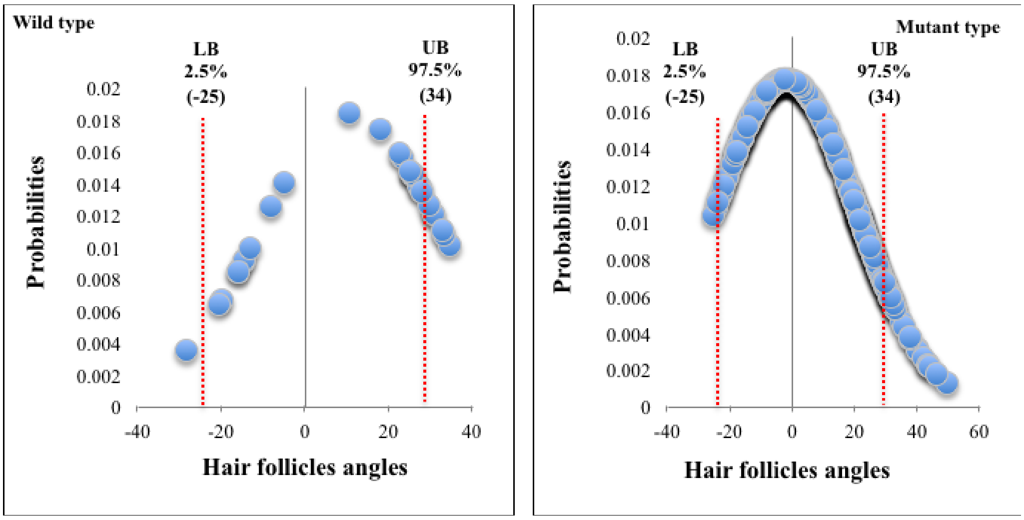


Figure 7.13 Normal distribution of median angles of the hair follicles for wild and mutant populations with RR bound, wild type on the left side and mutant on the right side

Figure 7.13 was showed the data distribution for the hair follicles angles in the epidermis layer into two types, wild (WT) and mutant. Lower Bound (LB) and the Upper Bound (UB) of the RR are determined based on wild type (-25 to 34), which are pointed as two vertical red lines in Figure 7.13. Then the interesting genes has determined based on LB and UB in the mutant type as shown in the right side in Figure 7.13 (see Chapter 6 Section 5 for more details).

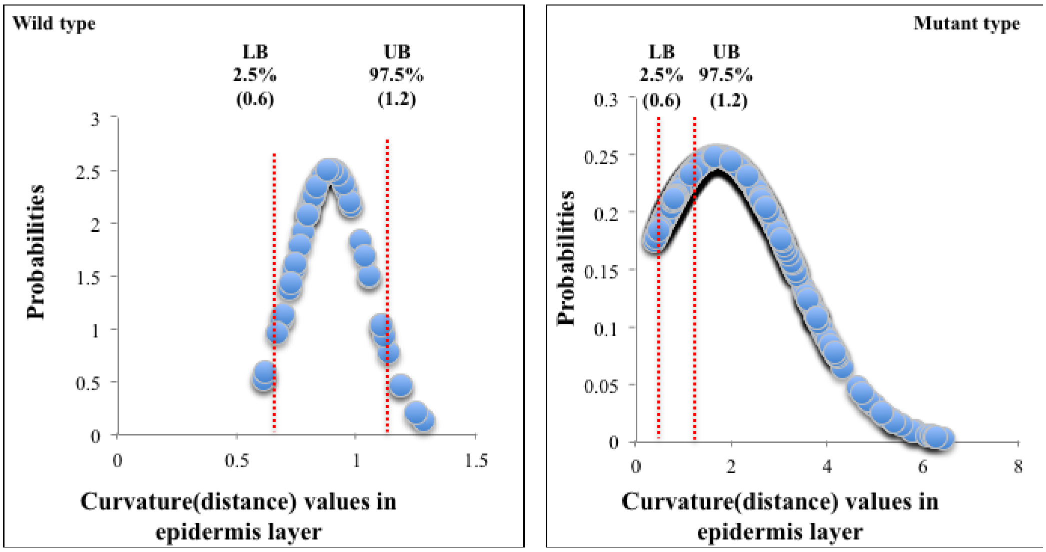


Figure 7.14 Normal distribution of the distances in the epidermis layer for wild and mutant populations with RR bound, wild type on the left side and mutant on the right side

Figure 7.14 was showed the data distribution for the distances of the epidermis curvature into two types, wild (WT) and mutant. Lower Bound (LB) and the Upper

Bound (UB) of the RR are determined based on wild type (0.6 to 1.2), which are pointed as two vertical red lines in Figure 7.14. Then the interesting genes has determined based on LB and UB in the mutant type as shown in the right side in Figure 7.14 (see Chapter 7 Section 4 for more details).

Considering the data distributions in the above Figures, we note that the best analysis to identify interesting genes was obtained with the analysis of nuclei counts in the dermis layer. This could be partially attributed to the fact that the dermis layer is rare to lose by damage sample even other layers are damaged in the same image (Figure 7.15). Compared to the epidermis parameters, this means we had more correctly segmented images to be analysed. Epidermis layer more likely to damage which lead to deficits samples for correct analysis in relation to hair follicle & curvature parameters (see Figures 7.16 and 7.17). In fact, we found that hair follicles are not available in all images due to damaged epidermis, or it doesn't exist i.e. deficits of samples.

Moreover the list of interesting genes discovered in the dermis layer (see Chapter 5 Section5) include a possible novel gene called ‘Traf3ip3/Traf3ip3’, which when knockout there would be increase in nuclei causing the tumour of melanoma cells (Nasarre *et al.*, 2018).

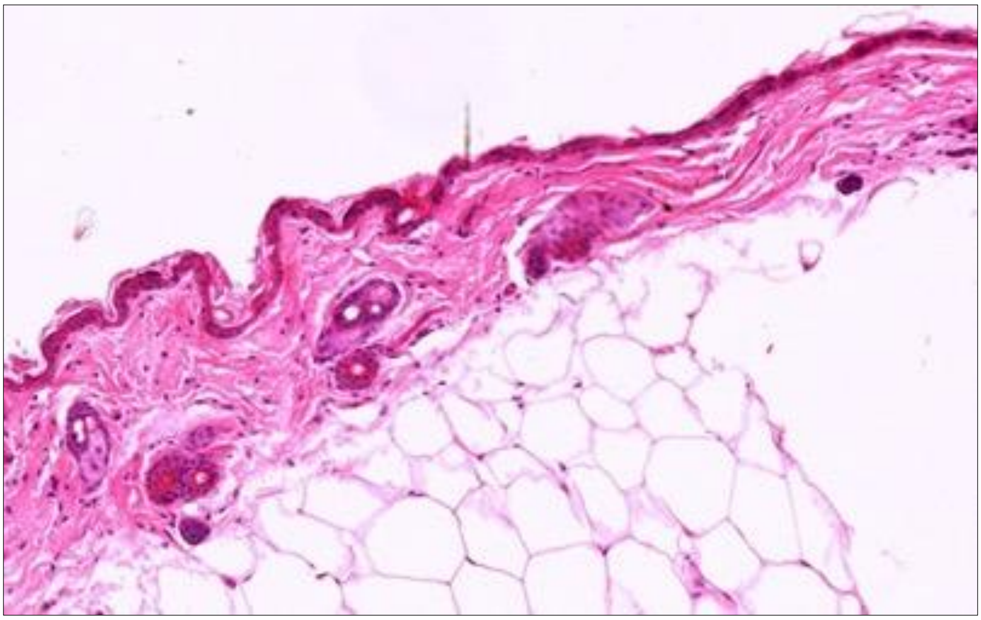


Figure 7.15 Correct samples of the dermis layer while damage in fat cells layer

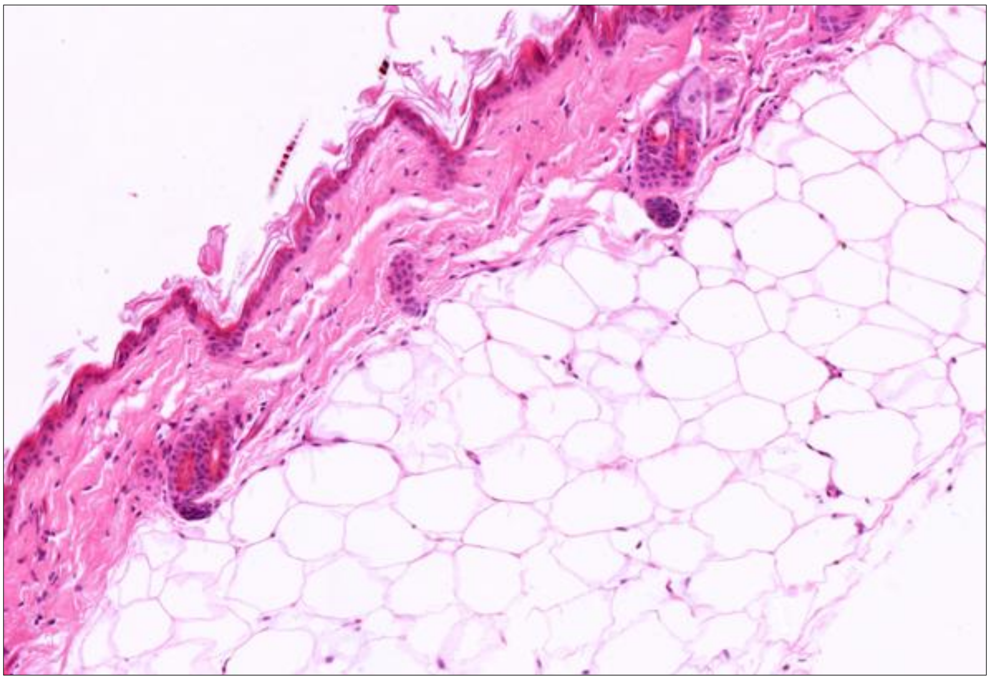


Figure 7.16 Samples without hair follicles

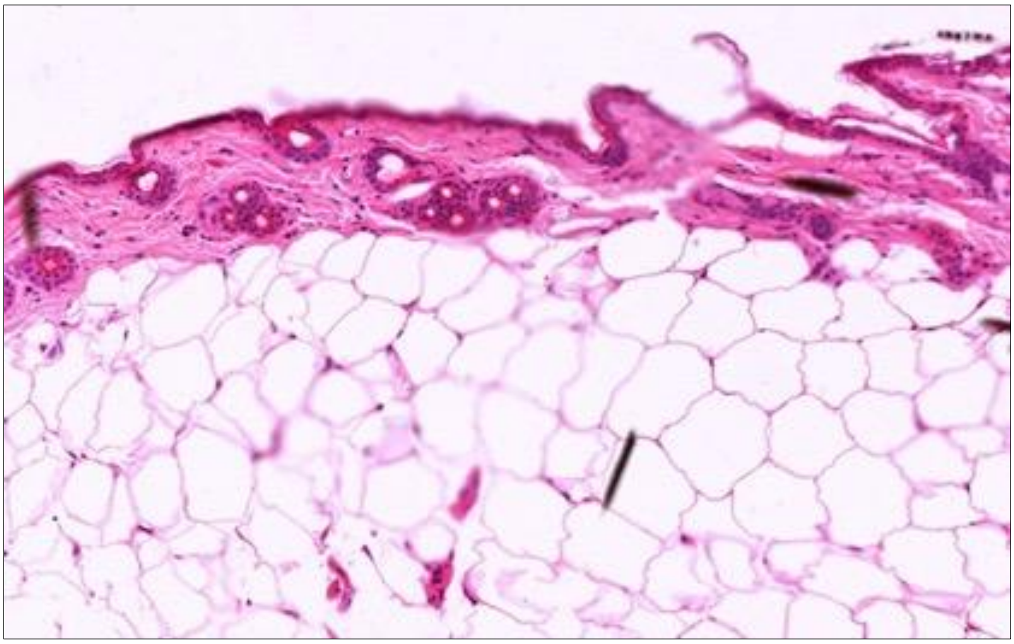


Figure 7.17 Damage in the epidermis layer

7.5 Conclusion

Biological image processing and analysis provide techniques that help scientists to evaluate the effects of physiological changes in a research context. These techniques may also be used to evaluate the effect of the treatments and drug efficiency in the context of the drug discovery. All of these analysis techniques could also be applicable clinically. In all applications, novel combinations of image processing/analysis techniques and pipelines will save time and are expected to produce more accurate results that will ultimately help or improve the speed and quality of dermatology and cosmetic treatments. This chapter described a successful method for quantifying the curvature of the epidermis to help identifying genes responsible for changes in the curvature and it was a successful method. Experiments on a large set of microscopic images of mutant and WT mice demonstrated the effectiveness of the proposal.

Chapter 8 : Conclusions and Future work

The specific aims and objectives for this thesis research project arose in the context of collaborative discussions between the then Buckingham Applied Computing department, the Buckingham Institute of Translation Medicine and researchers at the Sanger Institute, Cambridge, on automating certain image processing/analysis high throughput tasks conducted at the Sanger institute within their Mouse Genetics Project (MGP). The MGP is a large-scale mutant mice production and phenotyping initiative designed to integrate established knowledge of the mouse genome and of disease models.

Our specific area of research was targeting skin disease related phenotyping tasks, by analyzing their large dataset of H&E skin tissue microscopic images. Accordingly, our initial research work was focused and acquiring a working knowledge of the main features/objects in the H&E mouse skin tissue images, the parameters of which are to be investigated computationally for their causal effects on skin diseases. Mammalian skin tissue a complex structure of multiple layers containing different types of cells and objects that are visibly distinguishable under the microscope using well-known staining procedures. Dermatology specialists have long associated many diseases with deterioration of certain feature parameters associated with certain objects/cells visible in different skin tissue layers. The main focus of our research, was concerned with quantifying changes to the number of nuclei in the dermis and epidermis layers, the orientation of hair follicles, curvature of the outer border of the epidermis. Accordingly, the research work conducted through my PhD program was devoted to investigate, design, and develop reliably effective automatic image processing/analysis algorithms to quantify the above mentioned features/objects extracted from images of mouse skin tissue within their known layers. Furthermore, machine learning based experiments were conducted to identify the genetic causes of changes to these parameters in relation to skin diseases.

Automating the initial tasks of detecting the nuclei cells and the hair follicles presented the first challenge of preprocessing the images by segmenting the different skin layers and their sub-layers. Conventional threshold based and standard color based clustering proved to be of limited use in microscopic images due to a number of factors including potential staining errors resulting in overlaps, smugness, and/or missing parts. Inadequate fixed image magnification makes some of these tasks more complex to automate. Such a challenge necessitates the search for segmentation techniques that tolerate fussiness of separating borders of different layers/sub-layers. However, proving the effectiveness of any proposed scheme, in these tasks or the subsequent ones, rely greatly on knowing the ground truth on a sufficiently large and diverse set of H&E skin microscopic images. Overcoming such difficulties can only be overcome in close consultation with domain experts, i.e. adopting a multi-disciplinary research approach.

Once, the various skin layers are reliably segmented, i.e. the regions of interest are determined, biologically established knowledge of shapes and other characterizing factors become essential, but again overlapping of detected objects presented a non-obvious challenge that had to be dealt with appropriately. Moreover, interpreting and mimicking the way the trained life-science researchers or technicians determine the overlap of nuclei, the orientation of hair follicles, or epidermis curvature present added challenges. The fussy nature of the way the human make relevant decisions in these cases is only one aspect of the challenge. A more specific aspect of this type of challenges is how to appropriately model a computable mathematical function to quantify concepts like shape of a nuclei-like object or a hair follicle, or the curvature of the complex epidermis layer? In this respect, our developed schemes need to be able to take into effect staining errors on overlapping as well as the effect of unavoidable rapture of the epidermis, in different places, on curvature quantification as well as hair follicle orientation.

We developed appropriate computation schemes that deal with all the above challenges, and demonstrated their effectiveness by experiments on sufficiently large set of the images to show high accuracy agreement with the ground truth established manually by an expert life science research scientist. Using each of the developed schemes with over 5000 MGP generated H&E skin mouse images, at the Sanger Institute, and appropriate classification algorithms resulted in identifying genes that cause the relevant skin abnormalities. Thereby, the work carried achieved the main originally set objectives with high accuracy.

In short, this thesis demonstrated the utility of combining computational non-machine learning methods and unsupervised machine learning of image processing techniques. And the use of the automated high-throughput image analysis tools that are designed primarily to detect, segment and quantify various components of skin images. Then extracting the interesting changes in the skin object such as change in the nuclei numbers, hair follicle orientation and morphological feature to the nuclei in the skin.

8.2. Novel Contributions

In this section we shall first give details of the main novel schemes developed in each of the chapters beyond the introduction and background chapters. We then highlight the significance of the work conducted in thesis.

8.2.1 Image Preprocessing Tasks

Chapter 4 focused on extracting/segmenting the skin layers that are epidermis, dermis and subcutaneous layer. Our results led to a successful design of techniques that combine fuzzy c-mean cluster and color deconvolution, to segment the three main layer of mice skin and sub segment the epidermis layer in two layers basal and cornified. This automatic method, exploits the number of peaks in the histogram of the red channel of H&E skin images to determine appropriate mathematical morphology operations for accurate segmentation. The level of success was established experimentally by comparing the automatic segmentation with the ground truth (as determined by a domain expert researcher) on a sample of 100 set of the H&E skin images.

8.2.2 Quantification of skin objects/features for phenotype detection

Chapters 5, 6 and 7, reports the results of the investigations conducted for the development of a high-throughput automated image analysis system to detect interesting skin disease related cutaneous phenotypes in gene-knockout mice provided by WTSI.

In Chapter 5, nuclei in the epidermis layer and the dermis layer were segmented with high accuracy, when compared with the ground truth. Fuzzy clustering with mathematical morphology operations have shown to overcome the staining-related challenges faced by conventional methods. The overlap nuclei in the epidermis layer was separated as well. Using watershed transform with distance function was sufficient to deal with the problem of overlapped epidermis nuclei and apply adequate separation. Improved nuclei segmentation in the dermis layer benefit from using biologically established facts. The novelty in this improvement, is the using two filters to identify candidate overlap nuclei before separating them. The first filter base on the mean of two features, the circularity and aspect ratio, and the second filter been used the output image of the first filter, and then the mean of three features (circularity, major axis and perimeter) was used to extract the final candidate overlap nuclei.

In Chapter 6, was dedicated to the development of automatic methods for segmenting hair follicles in the epidermis layer, aligning the epidermis layer, and quantifying the orientation of the hair follicles against the aligned epidermis in relation to certain skin abnormalities. Orientation of a hair follicle, turned out to depend on the shape of the follicle which in turn depend on the accuracy of its segmentation. The shape of an accurately segmented follicle can be approximated by an elongated pocket where the two furthest away point can be used to determine its orientation. Problems with segmentation of the epidermis, damage to the epidermis were overcome by using the median of the orientation of the segmented follicles was used to determine the overall regional tissue follicles orientation.

In Chapter 7, the initial challenge was the need for a mathematical model for the curvature of the epidermis layer to reflect the way the human expert interpret curvature of a tissue of such complexity. Naturally, the way a tissue is curled locally around itself throughout the entire tissue is an indicator of the curvature value. Realizing that the actual area of a surface compared to the area of a plane that cuts through the surface can indirectly represent the curvature. Thus, we selected a reference line between the first point and end point in the top curve of the epidermis mask. Then quantification applied by measure the distance between each point in the curve and the reference line by the well-known mathematical equation representing the distance between a line and a point. Again the comparison with the ground truth as measured by the domain expert demonstrated that this model is reasonably reflective of the manual practice. This was confirmed by the experiments that used changes in the curvature in knockout gene tissue image to identify genes that could cause skin abnormalities.

8.2.3 The Significance of this Work

This approach is naturally applicable to the evaluation of the effect of treatments and drug efficiency in the context of the drug discovery. Such analytical techniques could also be integrated into clinical systems diagnosis and treatment assessments. In all applications, novel combinations of image processing/analysis techniques and pipelines will save time and are expected to produce more accurate results that will ultimately help or improve the speed and quality of dermatology and cosmetic treatments.

The high-throughput analysis application of our developed techniques is particularly evident in all chapters. However all the techniques can be utilized in this manner because of their automated and adaptive nature. The results from using the image analysis and quantification techniques have proved that these methods are reliable and accurate. Moreover, the techniques are easy to use and utilize for this type of translational research.

8.3 Applicability to the Human Images and other Mouse Data

All the works described in this thesis was designated to segment and annotate different region in the H&E images of the mouse skin, and extract some useful information from them. The direct and indirect benefits of the work achieved by the thesis as well as the impacts of our work can be summarized as follows:

1. Helping the identification of novel genes related to specific diseases such as skin cancer or other abnormalities. Even any defined gene is obtained from manipulation of the mouse dataset, should be highly relevant to the human skin diseases because of the significant similarity between the human and the mouse genetic backgrounds.
2. Providing Life Science researchers with consistent and accurate software tools to segment, annotate and quantify parameters of different mouse’s skin regions/objects.
3. Our layer segmentation solutions can be applied on the human H&E images for assisting researchers in accurate diagnoses, as we applied on the human image as shown result in Figure 8.1.

   To provide some evidence for this claim, we investigated the initial applicability of two of the proposed methods in this thesis to the microscopic human images after some adjustments. Results of applying the layer segmentation method on H&E image of humans can be seen in Figure 8.1. Our layer segmentation methods seem to provide good solutions for human H&E images especially when the skin layers are very difficult to be identified. Furthermore, the main steps of the proposed nuclei segmentation method were applied on the human H&E image with skin cancer and the result is shown in Figure 8.2. The result indicates that the nuclei segmentation method is indeed applicable to a human H&E image. With some adjustment, the nuclei can not only be detected and counted, but also segment the tumor (group of nuclei). We note that due to unavailable access to a sufficiently number of human H&E images, only one image is used in this case as a proof of concept.
4. Besides applicability of the corresponding parts on human medical images and different animal datasets, the algorithms and ideas developed here can be generalised to solve other problems in different microscopic imaging such as PAS

staining.

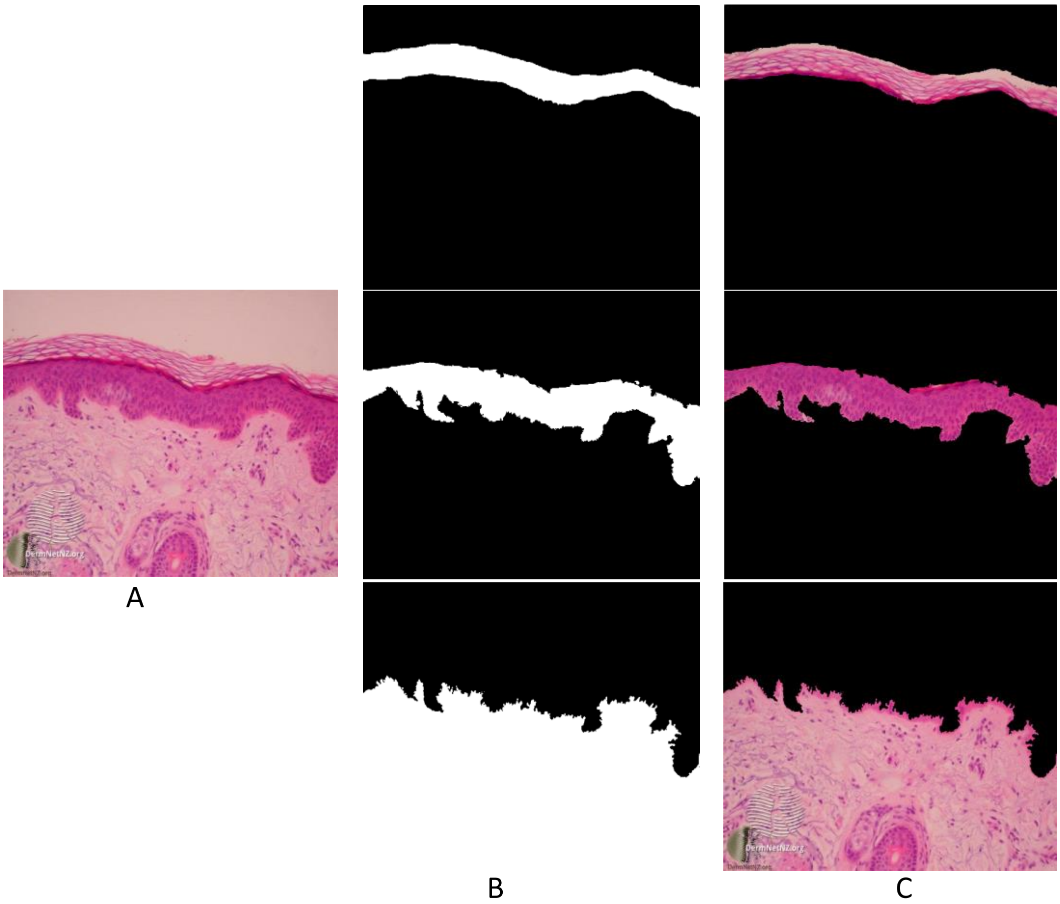


Figure 8.1 Human layer segmentation. A) Input image. B) Layers masks. C) Layers segmentation

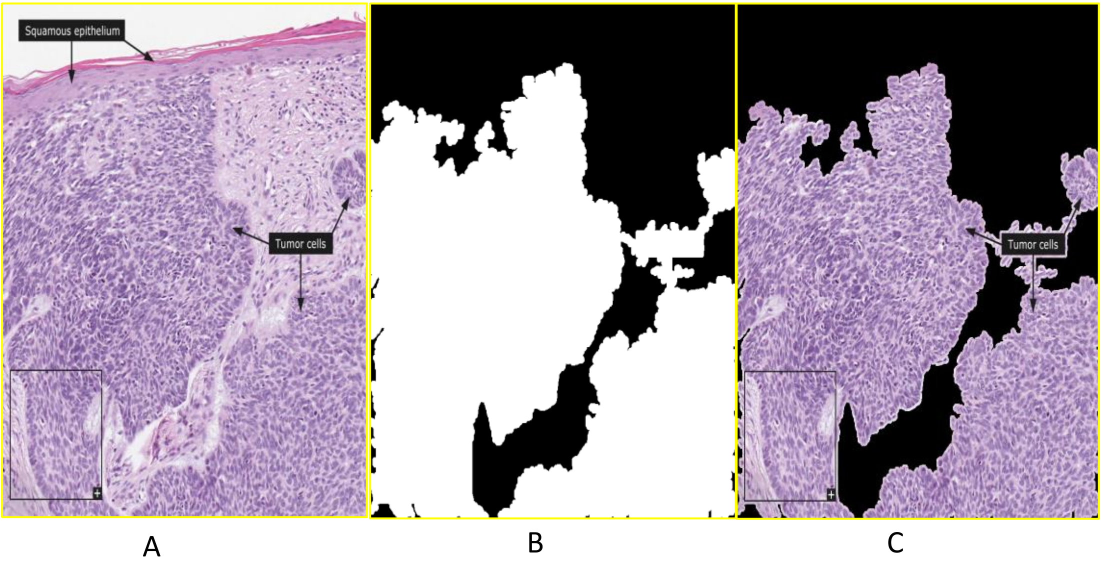


Figure 8.2 Nuclei segmentation (Tumour) from human H&E skin with cancer image

8.4 Future work

One future objectives include making the developed methods adaptive to different datasets. For example we shall attempt to develop appropriate image transformations that could be applied all the image dataset the same color distribution using color normalization (Reinhard *et al.*, 2001). This technique been used in the histopathological study to map the color distribution of bad stained images, e.g. over stained or under stained onto an image that exhibit characteristics of a well stained image (Magee *et al.*, 2009).

We would also like to contribute a method for automatic identifying the damage in skin layer (Schmitz *et al.*, 2018), such as damage in the dermis layer and damage in fat cells in subcutaneous layer as shown in Figure 8.3. The first step start with layer segmentation to separate the main skin layers as individual layer, which l have done in the chapter 4 by using a fusion method of layer segmentation. Next step need to use one of the texture classification methods based on texture feature. Then the texture will classify by using random forest method (Aygün, Yalç\in and Güne\cs, 2017). The method should be identify an image, which the damage layer, and ignore the image, because the image wasn't comparable with other undamaged images.

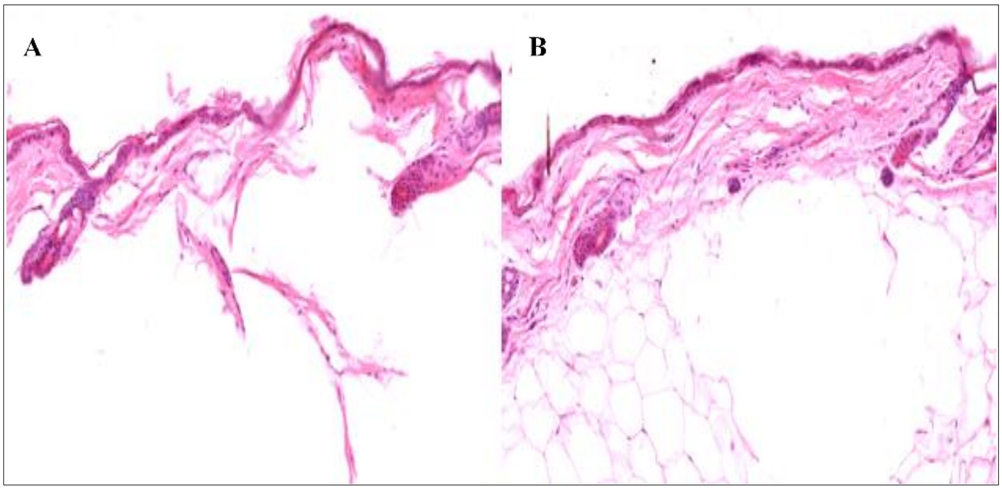


Figure 8.3 Example of layer damage. A) Damage in dermis layer. B) Damage in epidermis layer

In a future work, l would like to apply deep learning method on my development dataset to segment and quantify the nuclei in the epidermis and dermis layer. And l would like to compare the accuracy and time consume between the unsupervised techniques which I have already developed in chapter 4 and supervised method which it is deep learning method, and evaluate the best method of them based on results (Caicedo *et al.*, 2018)(Wollmann *et al.*, 2018).

As a future work for this research we would like to facilitate the use of various combination analysis tools to be used as a biological tool to assess the skin integrity. The graphical user interfaces (GUI) is important for the biologist to have a simple effort by following the GUI instruction, to get the analysis result quickly and gain an accurate result of analysis without bias that could affect the diagnosis. This can be used in the future as a part of WTSI pipeline to take a part in the phenotyping process, and also can be used as plug in for free analysis sources as imageJ to be available and accessed in biological image analysis. It would also be useful for training purposes.

Finally, a combination of image processing and analysis techniques can be used automatically in the high-throughput analysis to evaluate skin conditions in different diseases such as skin cancer, inflammatory disorders and cosmetic dermatology. The analyses in this field can be continued by developing an image analysis source to help solve concerning issues of time consumption and inaccurate assessments. And also there are opportunities for using such system in other application of medical image analysis e.g. ultrasound, magnetic resonance image MRI, x-ray images by modifying these techniques to help the researchers in this field (Holland and Marchand, 2002).

References

Adissu, H. A. *et al.* (2014) ‘Histopathology reveals correlative and unique phenotypes in a high-throughput mouse phenotyping screen’, *Disease Models & Mechanisms*. doi: 10.1242/dmm.015263.

Ahtiainen, L. *et al.* (2014) ‘Directional cell migration, but not proliferation, drives hair placode morphogenesis’, *Developmental cell*. Elsevier, 28(5), pp. 588–602.

Arvis, V. *et al.* (2011) ‘Generalization of the cooccurrence matrix for colour images: application to colour texture classification’, *Image Analysis & Stereology*, 23(1), pp. 63–72.

Awgulewitsch, A. (2003) ‘Hox in hair growth and development’, *Naturwissenschaften*. Springer, 90(5), pp. 193–211.

Aygün, S., Yalç\in, H. and Güne\cs, E. O. (2017) ‘Seed texture classification by random forest and neural networks’, in *Signal Processing and Communications Applications Conference (SIU), 2017 25th*, pp. 1–4.

Babu, M. N. *et al.* (2010) ‘Histo-pathological image analysis using os-fcm and level sets’, in *Applied Imagery Pattern Recognition Workshop (AIPR), 2010 IEEE 39th*, pp. 1–8.

Baecker, V. (2010) ‘Image processing and analysis with Image J and MRI Cell Image Analyzer’, *Montpellier RIO Imaging*, pp. 1–93.

Bancroft, J. D. and Gamble, M. (2008) *Theory and practice of histological techniques*. Elsevier Health Sciences.

Bapure, K. (2012) *Automated Image Analysis for Nuclear Morphometry Using H&E and Feulgen Stains in Prostate Biopsies*.

Barreto, J., Tita, D. and Orlandi, M. (2019) ‘Development of an Automated Method To Perform a Quantitative Study of Particle Size Distribution and the Effect of a Conductive Layer in Scanning Electron Microscopy’, *Química Nova*. doi: 10.21577/0100-4042.20170353.

Bayramoglu, N. *et al.* (2017) ‘Towards Virtual H&E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks.’, in *ICCV Workshops*, pp. 64–71.

Bhattacharyya, S. *et al.* (2019) *Hybrid Computational Intelligence: Research and Applications*. CRC Press. Available at: https://books.google.co.uk/books?id=m8O2DwAAQBAJ.

Bigby, M. (2011) *Atlas of Normal Human Skin, Archives of Dermatology*. Springer Science & Business Media. doi: 10.1001/archderm.1993.01680270148033.

Bland, J. M. and Altman, D. G. (2010) ‘Statistical methods for assessing agreement between two methods of clinical measurement’, *International Journal of Nursing Studies*. Elsevier, 47(8), pp. 931–936.

Brandt, A. *et al.* (2006) ‘Developmental control of nuclear size and shape by kugelkern and kurzkern’, *Current Biology*. Elsevier, 16(6), pp. 543–552. doi: 10.1016/j.cub.2006.01.051.

Britannica, E. (2018) *Adipose cell, Encyclopædia Britannica, inc.* Available at: https://www.britannica.com/science/adipose-cell (Accessed: 7 July 2019).

Brommage, R., Powell, D. R. and Vogel, P. (2019) ‘Predicting human disease mutations and identifying drug targets from mouse gene knockout phenotyping campaigns’, *DMM Disease Models and Mechanisms*. doi: 10.1242/dmm.038224.

Brüne, M. and Schiefenhövel, W. (2019) *Oxford Handbook of Evolutionary Medicine*. OUP Oxford. Available at: https://books.google.co.uk/books?id=ZviFDwAAQBAJ.

Bulten, W. *et al.* (2019) ‘Epithelium segmentation using deep learning in H&E-stained

prostate specimens with immunohistochemistry as reference standard', *Scientific Reports*. doi: 10.1038/s41598-018-37257-4.

Burns, T. (2010) 'Rook's Textbook of Dermatology. Print and Online Package'. Wiley-Blackwell.

Caicedo, J. C. *et al.* (2018) 'Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images', *bioRxiv*. Cold Spring Harbor Laboratory, p. 335216.

Cetera, M. *et al.* (2018) 'Counter-rotational cell flows drive morphological and cell fate asymmetries in mammalian hair follicles', *Nature cell biology*. Nature Publishing Group, 20(5), p. 541.

Chai, T. and Draxler, R. R. (2014) 'Root mean square error (RMSE) or mean absolute error (MAE)?--Arguments against avoiding RMSE in the literature', *Geoscientific model development*. Copernicus GmbH, 7(3), pp. 1247–1250.

Chang, H. *et al.* (2015) 'Identification of Astrotactin2 as a Genetic Modifier That Regulates the Global Orientation of Mammalian Hair Follicles', *PLoS Genetics*. Public Library of Science, 11(9), p. e1005532. doi: 10.1371/journal.pgen.1005532.

Chang, H., Smallwood, Philip M., *et al.* (2016) 'The spatio-temporal domains of Frizzled6 action in planar polarity control of hair follicle orientation', *Developmental Biology*. doi: 10.1016/j.ydbio.2015.10.027.

Chang, H., Smallwood, Philip M, *et al.* (2016) 'The spatio-temporal domains of Frizzled6 action in planar polarity control of hair follicle orientation', *Developmental biology*. Elsevier, 409(1), pp. 181–193.

Chang, H. and Nathans, J. (2013) 'Responses of hair follicle--associated structures to loss of planar cell polarity signaling', *Proceedings of the National Academy of Sciences*. National Acad Sciences, p. 201301430.

Chen, C., Li, H. and Zhou, X. (2007) 'Automated segmentation of drosophila RNAi fluorescence cellular images using graph cuts', in *International Conference on Multimedia Modeling*, pp. 116–125.

Chen, Q. *et al.* (2012) 'Modified two-dimensional Otsu image segmentation algorithm and fast realisation', *IET Image Processing*. IET, 6(4), p. 426. doi: 10.1049/iet-ipr.2010.0078.

Cheng, J., Rajapakse, J. C. and others (2009) 'Segmentation of clustered nuclei with shape markers and marking function', *IEEE Transactions on Biomedical Engineering*. IEEE, 56(3), pp. 741–748.

Chong, J. X. *et al.* (2015) 'The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2015.06.009.

Cisneros, F. J. *et al.* (2011) 'Histology image segmentation', *International Journal of Information Technology and Management*, 5(1), pp. 67–76.

Collins, F. S. *et al.* (1998) 'New goals for the U.S. Human Genome Project: 1998-2003', *Science*. doi: 10.1126/science.282.5389.682.

Collins, T. J. (2007) '{\guillemotleft}ImageJ for microscopy{\guillemotright}. BioTechniques. 43 (1 Suppl): 25-30'.

Cortes, C. and Vapnik, V. (1995) 'Support-vector networks', *Machine learning*. Springer, 20(3), pp. 273–297.

Cristianini, N., Shawe-Taylor, J. and others (2000) *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Cuadros, M. *et al.* (2007) 'Identification of a proliferation signature related to survival in nodal peripheral T-cell lymphomas', *Journal of clinical oncology*. New York, NY: Grune & Stratton, c1983-, 25(22), pp. 3321–3329.

Das, D. K., Maiti, A. K. and Chakraborty, C. (2018) 'Automated identification of normoblast cell from human peripheral blood smear images', *Journal of microscopy*.

Wiley Online Library, 269(3), pp. 310–320.

Devenport, D. and Fuchs, E. (2008) 'Planar polarization in embryonic epidermis orchestrates global asymmetric morphogenesis of hair follicles', *Nature cell biology*. Nature Publishing Group, 10(11), p. 1257.

Dhara, S. R. R., Suryono, S. and Widodo, C. E. (2017) 'Determination of breast cancer area from mammography images using thresholding method', *Int. J. Innov. Res. Adv. Eng*, 4(06), pp. 80–83.

DiTommaso, T. *et al.* (2014) 'Identification of genes important for cutaneous function revealed by a large scale reverse genetic screen in the mouse', *PLoS genetics*. Public Library of Science, 10(10), p. e1004705.

Driskell, R. R. *et al.* (2014) 'Defining dermal adipose tissue', *Experimental Dermatology*. doi: 10.1111/exd.12450.

Du, H., Chen, X. and Xi, J. (2019) 'An improved background segmentation algorithm for fringe projection profilometry based on Otsu method', *Optics Communications*. doi: 10.1016/j.optcom.2019.06.044.

Edens, L. J. *et al.* (2013) 'Nuclear size regulation: From single cells to development and disease', *Trends in Cell Biology*. doi: 10.1016/j.tcb.2012.11.004.

Esfahlani, F. Z. and Sayama, H. (2018) 'A percolation-based thresholding method with applications in functional connectivity analysis', in *International Workshop on Complex Networks*, pp. 221–231.

Esteva, A. *et al.* (2019) 'A guide to deep learning in healthcare', *Nature Medicine*. doi: 10.1038/s41591-018-0316-z.

Farage, M. A. *et al.* (2010) 'Cutaneous effects and sensitive skin with incontinence in the aged', in *Textbook of aging skin*. Springer, pp. 663–671.

Feng, X. *et al.* (2019) 'Biophysical basis of skin cancer margin assessment using Raman spectroscopy', *Biomedical Optics Express*. doi: 10.1364/boe.10.000104.

Fisher, R. *et al.* (2005) 'HIPR2: Image processing learning resources'.

Ford, A. and Roberts, A. (1998) 'Colour space conversions', *Westminster University, London*, 1998, pp. 1–31.

Fouad, S. *et al.* (2017) 'Unsupervised morphological segmentation of tissue compartments in histopathological images', *PloS one*. Public Library of Science, 12(11), p. e0188717.

Frost, J. (2013) *Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit? - Adventures in Statistics | Minitab*, *The Minitab Blog*.

Gauthier, B. E. *et al.* (2019) 'Toxicologic Pathology Forum*: Opinion on Integrating Innovative Digital Pathology Tools in the Regulatory Framework.', *Toxicologic pathology*. United States, 47(4), pp. 436–443. doi: 10.1177/0192623319827485.

Gawkrodger, D. and Ardern-Jones, M. R. (2016) *Dermatology E-Book: An Illustrated Colour Text*. Elsevier Health Sciences.

Genes in Life | Genetics 101 (no date). Available at: http://www.genesinlife.org/genetics-101/why-learn-about-genetics [Accessed 3 12 2016] (Accessed: 30 October 2019).

Genetics Home Reference (2003) *Genetics Home Refernce, online*. Available at: https://ghr.nlm.nih.gov/primer/hgp/genome (Accessed: 7 September 2016).

Gherardi, A. (2008) *A skin surface characterization system based on capacitive image analysis*. alma.

Goldman, G. (2008) 'Wound closure materials and instruments', *Bolognia dermatology. 2nd ed. Philadelphia, PA: Mosby*, p. 2188.

Gonzalez, R., Woods, R. and Eddins, S. (2004) 'Digital Image Processing Using MATLAB: Pearson Prentice Hall', *Upper Saddle River, New Jersey*. Available at: https://scholar.google.ca/scholar?q=Gonzalez+Digital+Image+Processing+Using+MATLAB%2C+New+Jersey%2C+Pearson+Prentice+Hall%2C+2004.&btnG=&hl=en&as_s

dt=0%2C5#0.

Guo, N., Hawkins, C. and Nathans, J. (2004) ‘From The Cover: Frizzled6 controls hair patterning in mice’, *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0402802101.

Guven, M. and Cengizler, C. (2014) ‘Data cluster analysis-based classification of overlapping nuclei in Pap smear samples’, *Biomedical engineering online*. BioMed Central, 13(1), p. 159.

H., F. *et al.* (2016) *Towards the next-generation of cancer cell lines: Derivation of an organoid biobank, European Journal of Cancer*.

Haake, A. *et al.* (2001) ‘Structure and function of the skin: overview of the epidermis and dermis’, *The biology of the skin*. The Parthenon Publishing Group Inc., New York NY, 2001, pp. 19–45.

Hadshiew, I. M. *et al.* (2004) ‘Burden of hair loss: stress and the underestimated psychosocial impact of telogen effluvium and androgenetic alopecia’, *Journal of investigative dermatology*. Elsevier, 123(3), pp. 455–457.

Haggerty, J. M. *et al.* (2014) ‘Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin’, *BMC medical imaging*. BioMed Central, 14(1), p. 7.

Holland, O. T. and Marchand, P. (2002) *Graphics and GUIs with MATLAB*. Chapman and Hall/CRC.

Honari, G. (2017) ‘Skin structure and function’, in *Sensitive Skin Syndrome, Second Edition*. doi: 10.1201/9781315121048.

Honda, K. *et al.* (2007) ‘Epidermal hyperplasia and papillomatosis in mice with a keratinocyte-restricted deletion of csk’, *Carcinogenesis*. Oxford University Press, 28(10), pp. 2074–2081.

Hosea, S. P., Ranichandra, S. and Rajagopal, T. K. P. (2011) ‘Color Image Segmentation--An Approach’, *Color Image Segmentation--An Approach*, 2(3), p. 0.

Huang, W. (2008) ‘An Interactive Algorithm for Blurred Medical Image Segmentation Based on Curve Fitting’, in *Genetic and Evolutionary Computing, 2008. WGEC’08. Second International Conference on*, pp. 332–335.

Huber, M. D. and Gerace, L. (2007) ‘The size-wise nucleus: nuclear volume control in eukaryotes’, *The Journal of cell biology*. Rockefeller University Press, 179(4), pp. 583–584.

Hughes, C. *et al.* (2013) ‘Robust alignment of prostate histology slices with quantified accuracy’, *IEEE Transactions on Biomedical Engineering*. IEEE, 60(2), pp. 281–291.

IMPC (no date) *online*. Available at: http://www.mousephenotype.org/.

Irshad, H. *et al.* (2014) ‘Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential’, *IEEE reviews in biomedical engineering*. IEEE, 7, pp. 97–114.

Jiang, P. *et al.* (2012) ‘Learning-based automatic breast tumor detection and segmentation in ultrasound images.’, in *ISBI*, pp. 1587–1590.

Khan, A. M. *et al.* (2014) ‘A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution’, *IEEE Transactions on Biomedical Engineering*. IEEE, 61(6), pp. 1729–1738.

Kimm, L. R. *et al.* (2007) ‘Frequent occurrence of deletions in primary mediastinal B-cell lymphoma’, *Genes, Chromosomes and Cancer*. Wiley Online Library, 46(12), pp. 1090–1097. doi: 10.1002/gcc.20495.

Koch, M. *et al.* (2011) ‘Fully automatic segmentation of wrist bones for arthritis patients’, in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pp. 636–640.

Koo, H. J. *et al.* (2017) ‘Quantitative Computed Tomography Features for Predicting Tumor Recurrence in Patients with Surgically Resected Adenocarcinoma of the Lung’, *PloS one*. Public Library of Science, 12(1), p. e0167955.

Kueh, H. Y. *et al.* (2008) ‘Image analysis for biology’.

Kurugol, S. *et al.* (2011) ‘Semi-automated algorithm for localization of dermal/epidermal junction in reflectance confocal microscopy images of human skin’, in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XVIII*, p. 79041A.

Larson, K. *et al.* (2011) ‘Hematoxylin and eosin tissue stain in Mohs micrographic surgery: A review’, *Dermatologic Surgery*. doi: 10.1111/j.1524-4725.2011.02051.x.

Li, D. *et al.* (2013) ‘Integrating a statistical background-foreground extraction algorithm and SVM classifier for pedestrian detection and tracking’, in *Integrated Computer-Aided Engineering*. doi: 10.3233/ICA-130428.

Liakath-Ali, K. *et al.* (2014) ‘Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen’, *Nature communications*. Nature Publishing Group, 5, p. 3540.

Libby, J. (2017) *Math for Real Life: Teaching Practical Uses for Algebra, Geometry and Trigonometry*. McFarland. Available at: https://books.google.co.uk/books?hl=en&lr=&id=imsJDgAAQBAJ&oi=fnd&pg=PA1&dq=J.+Libby,+Teaching+Practical+Uses+for+Algebra,+Geometry+and+Trigonometry,+McFarland+%26+Company,+2017.&ots=5_WoZyKV55&sig=F0bOLSwNYYArekFofBNO2N_6x2M#v=onepage&q=J. Libby%2C Teach.

Lu, C. and Mandal, M. (2012) ‘Automated segmentation and analysis of the epidermis area in skin histopathological images’, in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 5355–5359.

Lu, C. and Mandal, M. (2015) ‘Automated analysis and diagnosis of skin melanoma on whole slide histopathological images’, *Pattern Recognition*. Elsevier, 48(8), pp. 2738–2750.

Magee, D. *et al.* (2009) ‘Colour normalisation in digital histopathology images’, in *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*.

Malik, K. *et al.* (2019) ‘Ichthyosis molecular fingerprinting shows profound T H 17 skewing and a unique barrier genomic signature’, *Journal of Allergy and Clinical Immunology*. doi: 10.1016/j.jaci.2018.03.021.

Mandal, R., Gupta, M. and Kar, C. (2016) ‘Automated ROI Detection for Histological Image using Fuzzy C-means and K-means algorithm’, in *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on*, pp. 1173–1178.

Mann, C. J. *et al.* (2005) ‘High-resolution quantitative phase-contrast microscopy by digital holography’, *Optics Express*. Optical Society of America, 13(22), pp. 8693–8698.

Mao, K. Z., Zhao, P. and Tan, P.-H. (2006) ‘Supervised learning-based cell image segmentation for p53 immunohistochemistry’, *IEEE Transactions on Biomedical Engineering*. IEEE, 53(6), pp. 1153–1163.

Masson-Lecomte, A. *et al.* (2019) ‘CD8+ Cytotoxic Immune Infiltrate in Non-Muscle Invasive Bladder Cancer: A Standardized Methodology to Study Association with Clinico-Pathological Features and Prognosis’, *Bladder Cancer*. doi: 10.3233/BLC-180206.

Maurer, C. R., Qi, R. and Raghavan, V. (2003) ‘A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. IEEE, 25(2), pp. 265–270.

McConnell, A. M. *et al.* (2019) ‘Neural crest state activation in NRAS driven melanoma, but not in NRAS-driven melanocyte expansion’, *Developmental Biology*. doi: 10.1016/j.ydbio.2018.05.026.

Meijering, E. and van Cappellen, G. (2006) ‘Biological Image Analysis Primer’, *Erasmus MC, Rotterdam*. Citeseer.

Melcer, S. and Gruenbaum, Y. (2006) ‘Nuclear morphology: when round kernels do the Charleston’, *Current biology*. Elsevier, 16(6), pp. R195--R197.

Mokhtari, M. *et al.* (2014) ‘Computer aided measurement of melanoma depth of invasion in microscopic images’, *Micron*. Elsevier, 61, pp. 40–48. doi: 10.1016/j.micron.2014.02.001.

Molin, J., Thorstenson, S. and Lundström, C. (2014) ‘Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013’, *Journal of Pathology Informatics*. doi: 10.4103/2153-3539.129452.

Mouelhi, A., Sayadi, M. and Fnaiech, F. (2013) ‘Hybrid segmentation of breast cancer cell images using a new fuzzy active contour model and an enhanced watershed method’, in *Control, Decision and Information Technologies (CoDIT), 2013 International Conference on*, pp. 382–387.

Mustra, M., Grgic, M. and Zovko-Cihlar, B. (2014) ‘Alignment of X-ray bone images’, in *Telecommunications (BIHTEL), 2014 X International Symposium on*, pp. 1–4.

Muthukannan, K. and M, M. M. (2010) ‘Color Image Segmentation using K-means Clustering and Optimal Fuzzy C-Means Clustering’, in *Computational Intelligence*, pp. 229–234.

N, S. and S, V. (2016) ‘Image Segmentation By Using Thresholding Techniques For Medical Images’, *Computer Science & Engineering: An International Journal*. doi: 10.5121/cseij.2016.6101.

Nakamura, M. *et al.* (2013) ‘Mutant laboratory mice with abnormalities in hair follicle morphogenesis, cycling, and/or structure: an update’, *Journal of dermatological science*. Elsevier, 69(1), pp. 6–29.

Namburete, A. I. L., Rana, M. and Wakeling, J. M. (2011) ‘Computational methods for quantifying in vivo muscle fascicle curvature from ultrasound images’, *Journal of biomechanics*. Elsevier, 44(14), pp. 2538–2543.

Nasarre, P. *et al.* (2018) ‘TRAF3-interacting protein 3, a new oncotarget, promotes tumor growth in melanoma’, *Melanoma research*. Wolters Kluwer, 28(3), pp. 185–194.

National Human Genome Researche Institute (2015) *National Human Genome Researche Institute, online*. Available at: https://www.genome.gov/12514551/knockout-mice-fact-sheet/.

Nava, V. E. and Jaffe, E. S. (2005) ‘The pathology of NK-cell lymphomas and leukemias’, *Advances in anatomic pathology*. LWW, 12(1), pp. 27–34. doi: 10.1097/01.pap.0000151318.34752.80.

Neal, B. (2017) *Measuring Shape*. Edited by Reprint. CRC Pr I Llc.

Ng, C. R. *et al.* (2015) ‘Preliminary brain region segmentation using FCM and graph cut for CT scan images’, in *BioSignal Analysis, Processing and Systems (ICBAPS), 2015 International Conference on*, pp. 52–56.

Noroozi, N. and Zakerolhosseini, A. (2015) ‘Computerized measurement of melanocytic tumor depth in skin histopathological images’, *Micron*. Elsevier, 77, pp. 44–56.

Orlov, N. V *et al.* (2010) ‘Automatic classification of lymphoma images with transform-based global features’, *IEEE Transactions on Information Technology in Biomedicine*. IEEE, 14(4), pp. 1003–1013.

Osman, O. S. *et al.* (2013) ‘A novel automated image analysis method for accurate adipocyte quantification’, *Adipocyte*. doi: 10.4161/adip.24652.

Ovaere, P. *et al.* (2009) ‘The emerging roles of serine protease cascades in the epidermis’, *Trends in biochemical sciences*. Elsevier, 34(9), pp. 453–463.

Panagiotara, A. *et al.* (2019) ‘Implementation of Genomic Medicine’, in *Emery and Rimoin’s Principles and Practice of Medical Genetics and Genomics*. doi: 10.1016/b978-0-12-812536-6.00015-8.

Parvin, B. *et al.* (2007) ‘Iterative voting for inference of structural saliency and characterization of subcellular events', *IEEE Transactions on Image Processing*. IEEE, 16(3), pp. 615–623.

Paus, R. and Cotsarelis, G. (1999) 'The biology of hair follicles', *New England journal of medicine*. Mass Medical Soc, 341(7), pp. 491–497.

Peters, E. L. *et al.* (2019) 'Meijer and Vloedman's histochemical demonstration of mitochondrial coupling obeys Lambert–Beer's law in the myocardium', *Histochemistry and Cell Biology*. doi: 10.1007/s00418-018-1716-3.

Plotczyk, M. and Higgins, C. A. (2019) 'Skin biology', in *Biomaterials for Skin Repair and Regeneration*. doi: 10.1016/b978-0-08-102546-8.00001-7.

Prajapati, S. J. and Jadhav, K. R. (2015) 'Brain Tumor Detection By Various Image Segmentation Techniques With Introducation To Non Negative Matrix Factorization', *Ijarcce*, 4(3), pp. 599–603. doi: 10.17148/ijarcce.2015.43144.

Praveen, G. B. and Agrawal, A. (2015) 'Hybrid approach for brain tumor detection and classification in magnetic resonance images', in *Communication, Control and Intelligent Systems (CCIS), 2015*, pp. 162–166.

Qi, X. *et al.* (2012) 'Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set', *IEEE Transactions on Biomedical Engineering*. IEEE, 59(3), pp. 754–765.

R. Dougherty, E. and Shmulevich, I. (2012) 'On the Limitations of Biological Knowledge', *Current Genomics*. doi: 10.2174/138920212803251445.

Rädler, J. O. *et al.* (2019) 'Nanometer-scale colocalization microscopy of Streptococcus pneumoniae filaments', in. doi: 10.1117/12.2509927.

Ravni, A. *et al.* (2009) 'Planar cell polarity cadherin Celsr1 regulates skin hair patterning in the mouse.', *The Journal of investigative dermatology*, 129(10), p. 2507.

Reinhard, E. *et al.* (2001) 'Color transfer between images', *IEEE Computer Graphics and Applications*. IEEE, 21(5), pp. 34–41. doi: 10.1109/38.946629.

Rich, L. and Whittaker, P. (2005) 'Collagen and Picrosirius Red Staining : a Polarized Light Assessment of Fibrillar Hue and Spatial Distribution', *Brazilian Journal Morphologic Science*, 22(2), pp. 97–104. doi: 10.1016/j.eurpolymj.2004.01.020.

Rittscher, J., Machiraju, R. and Wong, S. T. C. (2008) *Microscopic image analysis for life science applications*. Artech House.

Rogojanu, R. *et al.* (2010) 'Segmentation of cell nuclei within complex configurations in images with colon sections', in *Intelligent Computer Communication and Processing (ICCP), 2010 IEEE International Conference on*, pp. 243–246.

Ruifrok, A. C., Johnston, D. A. and others (2001) 'Quantification of histochemical staining by color deconvolution', *Analytical and quantitative cytology and histology*. St. Louis, MO: Science Printers and Publishers, 1985-, 23(4), pp. 291–299.

Russ, J. C. and Neal, F. B. (2016) *The image processing handbook: Seventh edition, The Image Processing Handbook: Seventh Edition*.

S.C., T. (2002) 'Skin of color: Biology, structure, function, and implications for dermatologic disease', *Journal of the American Academy of Dermatology*. Elsevier, 46(2 III), pp. S41–S62. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=2005364507.

Sadeghian, F. *et al.* (2009) 'A framework for white blood cell segmentation in microscopic blood images using digital image processing', *Biological procedures online*. BioMed Central, 11(1), p. 196.

Sandby-Moller, J., Poulsen, T. and Wulf, H. C. (2003) 'Epidermal thickness at different body sites: relationship to age, gender, pigmentation, blood content, skin type and smoking habits', *Acta Dermato Venereologica*. Stockholm, Society for the Publication of Acta Dermato-Venereologica; distributed by Scandinavian University Press, etc., 83(6), pp. 410–413.

Sanz-Gómez, N., Freije, A. and Gandarillas, A. (2019) ‘Keratinocyte Differentiation by Flow Cytometry’, in. doi: 10.1007/7651_2019_237.

Sauermann, K. *et al.* (2002) ‘Age related changes of human skin investigated with histometric measurements by confocal laser scanning microscopy in vivo’, *Skin research and Technology*. Wiley Online Library, 8(1), pp. 52–56.

von Scheidt, M. *et al.* (2017) ‘Applications and Limitations of Mouse Models for Understanding Human Atherosclerosis.’, *Cell metabolism*. Elsevier Inc., 25(2), pp. 248–261. doi: 10.1016/j.cmet.2016.11.001.

Schmitz, L. *et al.* (2018) ‘Histological findings after argon plasma coagulation: an ex-vivo study revealing a possible role in superficial ablative treatment of the skin’, *Archives of dermatological research*. Springer, 310(2), pp. 157–163.

Schneider, M. R., Schmidt-Ullrich, R. and Paus, R. (2009) ‘The hair follicle as a dynamic miniorgan’, *Current Biology*. Elsevier, 19(3), pp. R132--R142.

Scott, A. L. (2015) *The Influence of Surgical Stress on Human Scalp Hair Fiber Dimensions*. Environment: Department of Archaeology.

Shih, H.-C. (2015) ‘An unsupervised hair segmentation and counting system in microscopy images’, *IEEE Sensors Journal*. IEEE, 15(6), pp. 3565–3572.

Shuaib, I. L. *et al.* (2018) *FRCR physics MCQs in clinical radiology, FRCR Physics MCQs in Clinical Radiology*. doi: 10.1007/978-981-13-0911-3.

Sklansky, J. and Bisconte, J. C. (2013) *Biomedical Images and Computers: Selected Papers Presented at the United States-France Seminar on Biomedical Image Processing, St. Pierre de Chartreuse, France, May 27--31, 1980*. Springer Berlin Heidelberg (Lecture Notes in Medical Informatics). Available at: https://books.google.co.uk/books?id=-nbpCAAAQBAJ.

St-Charles, P.-L. and Bilodeau, G.-A. (2014) ‘Improving background subtraction using local binary similarity patterns’, in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pp. 509–515.

Stenn, K. S. and Paus, R. (2001) ‘Controls of hair follicle cycling’, *Physiological reviews*. American Physiological Society Bethesda, MD, 81(1), pp. 449–494.

Suganya, R. *et al.* (2018) ‘Big Data in Medical Image Processing’, in *Big Data in Medical Image Processing*. doi: 10.1201/b22456-1.

Sundberg, J. P. (1994) *Handbook of mouse mutations with skin and hair abnormalities: animal models and biomedical tools*. CRC Press.

Sundberg, J. P. and King, L. E. (1996) ‘Mouse mutations as animal models and biomedical tools for dermatological research’, *Journal of Investigative Dermatology*. doi: 10.1111/1523-1747.ep12343152.

Takahashi, S. *et al.* (2019) ‘Genome-wide stability of the DNA replication program in single mammalian cells’, *Nature Genetics*. doi: 10.1038/s41588-019-0347-5.

Toennies, K. D. (2017) *Guide to medical image analysis*. Springer.

van der Veen, C. *et al.* (1999) ‘A comprehensive guide for the recognition and classification of distinct stages of hair follicle morphogenesis’, *Journal of Investigative Dermatology*. Elsevier, 113(4), pp. 523–532.

de Vries, H. J. C. *et al.* (2000) ‘Dermal organization in scleroderma: the fast Fourier transform and the laser scatter method objectify fibrosis in nonlesional as well as lesional skin’, *Laboratory investigation*. Nature Publishing Group, 80(8), p. 1281.

Wallentin, L. and Lindahl, B. (2019) ‘Uppsala Clinical Research Center—development of a platform to promote national and international clinical science’, *Upsala Journal of Medical Sciences*. doi: 10.1080/03009734.2018.1540506.

Wang, Q. A. *et al.* (2013) ‘Tracking adipogenesis during white adipose tissue development, expansion and regeneration’, *Nature Medicine*. doi: 10.1038/nm.3324.

Wang, X.-N. *et al.* (2014) ‘A three-dimensional atlas of human dermal leukocytes, lymphatics, and blood vessels’, *Journal of Investigative Dermatology*. Elsevier, 134(4),

pp. 965–974.
Wang, Y. *et al.* (2009) ‘Assisted diagnosis of cervical intraepithelial neoplasia (CIN)’, *IEEE Journal of Selected Topics in Signal Processing*. IEEE, 3(1), pp. 112–121.
Wang, Y. *et al.* (2016) ‘Patterning of papillae on the mouse tongue: A system for the quantitative assessment of planar cell polarity signaling’, *Developmental Biology*. doi: 10.1016/j.ydbio.2016.09.004.
Wang, Y., Badea, T. and Nathans, J. (2006) ‘Order from disorder: Self-organization in mammalian hair patterning’, *Proceedings of the National Academy of Sciences*. National Acad Sciences, 103(52), pp. 19800–19805.
Wang, Y., Chang, H. and Nathans, J. (2010) ‘When whorls collide: the development of hair patterns in frizzled 6 mutant mice’, *Development*. Oxford University Press for The Company of Biologists Limited, 137(23), pp. 4091–4099.
Wei, B.-R. *et al.* (2019) ‘Agreement in Histological Assessment of Mitotic Activity Between Microscopy and Digital Whole Slide Images Informs Conversion for Clinical Diagnosis’, *Academic Pathology*. doi: 10.1177/2374289519859841.
Weir, D. *et al.* (2012) ‘A user-specific machine learning approach for improving touch accuracy on mobile devices’, in *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 465–476.
Wellcome trust Sanger Institute (2016) *Mouse Pipelines, online*. Available at: https://www.sanger.ac.uk/science/groups/mouse-pipelines (Accessed: 3 December 2016).
Wen, S. *et al.* (2017) ‘A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image’, *Journal of Pathology Informatics*. doi: 10.4103/jpi.jpi_43_17.
White, J. K., Gerdin, A.-K., Karp, N. A., Ryder, E., Buljan, M., Bussell, J. N., Salisbury, J., Clare, S., Ingham, N. J., Podrini, C. and others (2013) ‘Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes’, *Cell*. Elsevier, 154(2), pp. 452–464.
White, J. K., Gerdin, A.-K., Karp, N. A., Ryder, E., Buljan, M., Bussell, J. N., Salisbury, J., Clare, S., Ingham, N. J., Podrini, C., Houghton, R., *et al.* (2013) ‘Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes The Sanger Institute Mouse Genetics Project’, *CELL*. doi: 10.1016/j.cell.2013.06.022.
Wienert, S. *et al.* (2012) ‘Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach’, *Scientific reports*. Nature Publishing Group, 2, p. 503.
Win, K. Y., Choomchuay, S. and Hamamoto, K. (2017) ‘K mean clustering based automated segmentation of overlapping cell nuclei in pleural effusion cytology images’, in *Advanced Technologies for Communications (ATC), 2017 International Conference on*, pp. 265–269.
Wollmann, T. *et al.* (2018) ‘Multi-channel Deep Transfer Learning for Nuclei Segmentation in Glioblastoma Cell Tissue Images’, in *Bildverarbeitung für die Medizin 2018*. Springer, pp. 316–321.
Woodcock, C. L. and Ghosh, R. P. (2010) ‘Chromatin higher-order structure and dynamics’, *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Lab, p. a000596.
Wu, J. *et al.* (2014) ‘Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation’, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 256–263.
Xiong, G. *et al.* (2006) ‘Automated neurite labeling and analysis in fluorescence microscopy images’, *Cytometry Part A: The Journal of the International Society for Analytical Cytology*. Wiley Online Library, 69(6), pp. 494–505.

Xu, H. *et al.* (2017) ‘Automatic measurement of melanoma depth of invasion in skin histopathological images’, *Micron*. doi: 10.1016/j.micron.2017.03.004.

Xu, H. and Mandal, M. (2015a) ‘Efficient segmentation of skin epidermis in whole slide histopathological images’, in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 3855–3858.

Xu, H. and Mandal, M. (2015b) ‘Epidermis segmentation in skin histopathological images based on thickness measurement and k-means algorithm’, *Eurasip Journal on Image and Video Processing*. doi: 10.1186/s13640-015-0076-3.

Yamasaki, T. *et al.* (2012) ‘Comparative study of interactive seed generation for growcut-based fast 3d mri segmentation’, in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pp. 1–4.

Yambal, M. and Gupta, H. (2013) ‘Image Segmentation using Fuzzy C Means Clustering : A survey’, in *International Journal of Advanced Research in Computer and Communication Engineering*, pp. 2927–2929. Available at: www.ijarcce.com.

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

Yousef, H. and Sharma, S. (2018) *Anatomy, Skin (Integument), Epidermis, StatPearls*.

Zainuddin, R., Naji, S. and Al-Jaafar, J. (2010) ‘Suppressing false nagatives in skin segmentation’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-642-17569-5_15.

Zhang, L. *et al.* (2011) ‘A practical segmentation method for automated screening of cervical cytology’, in *Intelligent Computation and Bio-Medical Instrumentation (ICBMI), 2011 International Conference on*, pp. 140–143.

Zhang, X. *et al.* (2017) ‘An improved fuzzy algorithm for image segmentation using peak detection, spatial information and reallocation’, *Soft Computing*. Springer, 21(8), pp. 2165–2173.

Zhang, Y. *et al.* (2013) ‘Automatic self-alignment and registration for PET/CT reconstruction by a cross-correlation maximization method’, in *IEEE Nuclear Science Symposium Conference Record*, pp. 1–4. doi: 10.1109/NSSMIC.2013.6829384.

Zhu, S. *et al.* (2015) ‘Identification of cancerous gastric cells based on common features extracted from hyperspectral microscopic images’, *Biomedical Optics Express*. doi: 10.1364/boe.6.001135.