



Analysis and characterization of the tissue structure of the epidermis from confocal imaging

Imane Lboukili

► To cite this version:

Imane Lboukili. Analysis and characterization of the tissue structure of the epidermis from confocal imaging. Computer Vision and Pattern Recognition [cs.CV]. Université Côte d'Azur, 2023. English. NNT : 2023COAZ4067 . tel-04331256

HAL Id: tel-04331256

<https://theses.hal.science/tel-04331256>

Submitted on 8 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT

Analyse et caractérisation de la structure tissulaire de l'épiderme à partir d'imagerie confocale

Imane LBOUKILI

Laboratoire d'Informatique, de Signaux et Systèmes de Sophia Antipolis
et Institut National de Recherche en Informatique et en Automatique

**Présentée en vue de l'obtention
du grade de docteur en Automatique
traitement du signal et des images**

D'Université Côte d'Azur

Dirigée par : Xavier DESCOMBES, DR1
INRIA, CNRS, I3S, Université Côte d'Azur,
France

Co-encadrée par : Georgios STAMATAS,
Research Associate Director & Fellow,
Translational Science, Johnson & Johnson,
France

Soutenue le : 26/10/2023

Devant le jury, composé de :

Alin ACHIM, Professor of Computational
Imaging, University of Bristol, England
Christophe ZIMMER, Director of Research
(HDR), Institut Pasteur, France
Josiane ZERUBIA, Director of Research
(HDR) INRIA, Université Côte d'Azur,
France
Salvador GONZÁLEZ, Professor of
Medicine and Medical Specialties,
Universidad de Alcalá, Spain

Analyse et caractérisation de la structure tissulaire de l'épiderme à partir d'imagerie confocale

Analysis and characterization of the tissue structure of the epidermis from confocal imaging

Imane Lboukili



Jury :

Présidente du jury

Josiane ZERUBIA, Director of Research (HDR) INRIA, Université Côte d'Azur, France

Rapporteurs

Alin ACHIM, Professor of Computational Imaging, University of Bristol, England

Christophe ZIMMER, Director of Research (HDR), Institut Pasteur, France

Examineurs

Salvador GONZÁLEZ, Professor of Medicine and Medical Specialties, Universidad de Alcalá, Spain

Directeur de thèse

Xavier DESCOMBES, DR INRIA, CNRS, I3S, Université Côte d'Azur, France

Co-encadrant de thèse

Georgios STAMATAS, Ph.D., Research Associate Director & Fellow, Translational Science, Johnson & Johnson, France

Analyse et caractérisation de la structure tissulaire de l'épiderme à partir d'imagerie confocale

La segmentation précise des cellules de l'épiderme sur les images de microscopie à réflectance confocale est importante pour l'étude non invasive de la structure tissulaire de l'épiderme de la peau saine et malade. Ces travaux sont néanmoins limités par le nombre d'échantillons traités du fait du recours intensif à l'utilisateur qui doit notamment pointer manuellement chacune des cellules. L'objectif de ce travail doctoral est de développer et d'implémenter une chaîne de traitements automatique pour analyser la structure de la peau, dans différents contextes, à partir d'images confocales.

Notre première contribution est un pipeline d'analyse d'images qui repose sur la détection des kératinocytes en fonction des caractéristiques morphologiques des membranes les entourant, appelé FIAP. Tout d'abord, nous identifions la région d'intérêt contenant les cellules, puis nous identifions ces cellules individuellement au sein de la région d'intérêt à l'aide d'applications successives de filtres de Sato et de Gabor. La dernière étape est le post-traitement des résultats précédemment obtenus pour l'amélioration de la détection des cellules et la suppression des régions détectées de taille aberrante. Nous avons évalué le pipeline proposé sur des données réelles annotées manuellement et sur des données de synthèses préalablement générées. Nous avons appliqué le pipeline à 5345 images de la joue et de l'avant-bras de participants âgés de 3 mois à 80 ans pour étudier l'évolution de l'architecture de l'épiderme au cours de la maturation et du vieillissement de la peau, ce qui en fait la première étude à grande échelle d'images de microscopie confocale. Nous avons démontré que la peau mûrit dynamiquement pendant l'enfance et vieillit à l'âge adulte, car la taille des kératinocytes augmente avec l'âge sur la joue et l'avant-bras, tandis que la topologie et le rapport d'aspect cellulaire restent inchangés à travers différentes couches épidermiques, parties du corps, et âge.

La deuxième méthode développée, appelée DermoGAN, est une nouvelle architecture de réseau de neurones fondée sur l'exécution de deux tâches en parallèles, constituée chacune d'un *cycle generative adversarial network* (cycle-GAN). La première tâche apprend à traduire des images de microscopie à réflectance confocale en segmentations binaires, apprenant ainsi le modèle de bruit et de texture des images, tandis que la seconde tâche transforme les images de microscopie à réflectance confocale filtrées par un filtre de Gabor en segmentations binaires, apprenant ainsi la structure épidermique visible sur les images. Nous affinons la segmentation en appliquant l'algorithme StarDist pour détecter les formes convexes, refermant ainsi toutes les membranes incomplètes et séparant les cellules voisines. DermoGAN a également été validé sur des données réelles annotées manuellement et comparé au FIAP. Les deux méthodes proposées ont été comparées à des approches de *machine learning* fondées sur U-net, Cell Cutter et cycle-GAN.

Nous montrons que DermoGAN est polyvalent, adaptable et généralisable car il peut être utilisé sur des images générées par d'autres techniques d'imagerie de différents tissus et donne des résultats précis sans réentraînement du modèle. Nous concluons que la précision de DermoGAN dépend de l'apparence de l'architecture tissulaire dans l'ensemble de données d'entraînement (cellules confluentes par rapport aux cellules indépendantes) plutôt que du type d'images, de l'organe ou du tissu visibles sur les images.

Nous avons exploré la possibilité d'utiliser DermoGAN pour la traduction des images de microscopie à réflectance confocale en histologie.

Enfin, nous avons développé une interface utilisateur regroupant toutes les méthodes développées.

Mots-clés : microscopie à réflectance confocale, kératinocytes, analyse d'images, multi-tâches, cycle-GAN, traduction d'images non appariées.

Analysis and characterization of the tissue structure of the epidermis from confocal imaging

Accurate segmentation of epidermal cells on reflectance confocal microscopy (RCM) images is important in the non-invasive study of epidermal architecture and topology of both healthy and diseased skin. However, analysis of these images is currently done manually, and therefore time-consuming, subject to human error and subjective interpretation, while at the same time being hindered by low image quality due to noise and signal heterogeneity. In this thesis, we design and implement two methods for automatic cell segmentation of granular and spinous RCM images of the human epidermis.

Our first contribution is a full image analysis pipeline (FIAP) which relies on the detection of keratinocytes based on their membrane morphological features. First, we identify the region-of-interest (ROI) containing cells and we then identify individual cells within the ROI using successive applications of Sato and Gabor filters. The final step is a post-processing improvement of cell detection and removal of size outliers. We evaluated the proposed pipeline on manually annotated real data and generated synthetic data, and applied it to 5345 images of the cheek and volar forearm of participants, ages 3 months to 80 years, to study the evolution of epidermal architecture during skin maturation and ageing, making it the first large-scale study of RCM images and showing that skin dynamically matures during childhood and ages during adulthood, as keratinocyte size increases with age on both cheek and volar forearm, while topology and cell aspect ratio remain unchanged across different epidermal layers, body sites and age.

The second developed method, called DermoGAN, is a novel neural network architecture based on an unsupervised dual-task cycle generative adversarial network (cycle-GAN). The first task learns to translate real RCM images into binary segmentations thus learning the noise and texture model of RCM images, while the second task maps Gabor-filtered RCM images into binary segmentations, learning the epidermal structure on RCM images. We refine the segmentation by applying the pre-trained StarDist algorithm to detect star-convex shapes, thus closing any incomplete membranes and separating neighboring cells. DermoGAN was also validated on manually annotated real data and compared to the FIAP. Both proposed methods were compared to machine learning approaches based on U-net, Cell Cutter and cycle-GANs.

We show that DermoGAN is versatile, domain adaptable and generalizable as it can be used on images generated by other types of imaging techniques of different tissues and gives accurate results with no retraining of the model. We conclude that the accuracy of DermoGAN depends on the nature of the tissue architecture in the training dataset (confluent cells vs. independent cells) rather than the type of images or the precise organ or tissue.

We explore the possibility of using DermoGAN for RCM to histology translation to visualize RCM insights in the form of histology slides for users who are not familiar with RCM.

Finally, we develop a user interface where users can find all the developed methods.

Keywords: reflectance confocal microscopy, keratinocytes, image analysis, multi-task, cycle-GAN, unpaired image translation.

Acknowledgments

I would like to express my heartfelt appreciation to my supervisors, Xavier Descombes and Georgios Stamatas. I am immensely grateful for your guidance, mentorship, and support during these 3 years.

I am deeply appreciative to the members of my thesis committee, Alin Achim, Salvador Gonzalez, Josiane Zerubia, and Christophe Zimmer. Thank you for your insightful comments on my work, I am grateful for the time you have dedicated to providing feedback.

Many thanks to everyone at Johnson & Johnson Santé Beauté France, who made this thesis and its completion feasible.

I would also like to thank my colleagues both at INRIA and Johnson & Johnson. First off, Fanny Alsamad and Elea Thibaut-Greugny, for their support and advice, and for being the best office mates one can hope for. You have filled these last 3 years with pockets of fun and joy. I would also like to thank Eric Debreuve for his help and support on D-accuracy, which was essential to this work.

Lastly, my deepest gratitude goes to my friends and family, whose unwavering belief in me has been a constant source of strength. Your love and encouragement are essential to my happiness and success. To my mom and siblings, thank you for going above and beyond for me. To my dad, I wish you were still with us to witness my success. To Antoine, thank you for your love and support.

Thank you all for being part of this academic and life milestone.

Acronyms

Reflectance confocal microscopy	RCM
Full image analysis pipeline	FIAP
Generative adversarial network	GAN
Cycle generative adversarial network	cycle-GAN
Optical coherence tomography	OCT
Dermal epidermal junction	DEJ
<i>Stratum corneum</i>	SC
<i>Stratum granulosum</i>	SG
<i>Stratum spinosum</i>	SS
<i>Stratum basale</i>	SB
Supra-papillary epidermis	SPE
Trans-epidermal water loss	TEWL
Principal components analysis	PCA
Convolutional neural network	CNN
Marked point process	MPP
Fully convolutional neural network	fCNN
Region of interest	ROI
Support vector machine	SVM
Grey level co-occurrence matrix	GLCM
Graphical user interface	GUI
Line-field confocal optical coherence tomography	LC-OCT
Singular value decomposition	SVD

List of Figures

Fig. 1 (A) RCM stacks are gray-scale images, orthogonal to the vertical sections typical of skin histology. The left panel shows an illustration of the undulating DEJ in blue. The depth limit of RCM is delineated in green, and RCM optical sections are represented with black dotted lines. The right panel shows a stack of images at sequential depths corresponding to these optical sections. (B) Diagram of an *in vivo* reflectance confocal microscope. (C) Image of a VivaScope 1500 (Lucid, Inc., Rochester, New York) reflectance confocal microscope. RCM, Reflectance confocal microscopy; DEJ, Dermal epidermal junction.

Fig. 2 Representative RCM images of minimally pigmented skin acquired at depths corresponding to the (A) *stratum corneum*, (B) *stratum granulosum*, (C) *stratum spinosum* and (D) *stratum basale*. In red, examples of granular and spinous keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy. Contrast was adjusted for clearer images.

Fig. 3 Representative RCM images of a heavily pigmented skin acquired at depths corresponding to the (A) *stratum corneum*, (B) *stratum granulosum*, (C) *stratum spinosum* and (D) *stratum basale*. In red, examples of granular and spinous keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy.

Fig. 4 RCM images with bright spots due to (A) keratin in hair shafts and (B) clustered keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy.

Fig. 5 Diagram of epidermal layer classification used by Somoza *et al.*

Fig. 6 Results of applying different traditional methods for segmentation on a same image.

Fig. 7 U-net architecture. A teal box represents a feature map. The number of channels is indicated on top of the feature map box. The image size in pixels is indicated on the lower left side of the box. Boxes with dark blue contours represent copied feature maps. The arrows indicate the different operations.

Fig. 8 RCM image of the *stratum spinosum* of minimally pigmented skin, Fitzpatrick phototype II. In blue, the border between tissue and background formed by micro-relief lines. In pink, non-informative areas. In orange, bright spots. In red, epidermal cells. RCM, Reflectance confocal microscopy. Image contrast was adjusted for better visualization. Scale bar = 50 μm .

Fig. 9 Steps of the identification of the ROI. The borders between the tissue and the background are identified using a Snake algorithm. The ROI was then refined using a Support Vector Machine algorithm trained to detect the non-informative areas, and a succession of morphological operations to remove bright spots in the RCM image. In blue is the border of the region of interest. In pink is an area that should have been removed through the different ROI identification steps but was not. Image contrast was adjusted for easier visualization. Scale bar = 50 μm . ROI, region of interest; RCM, Reflectance confocal microscopy.

Fig. 10 Four Haralick features were used to train an SVM to distinguish between non-informative and informative areas in RCM images. **(A)** Contrast vs. Homogeneity per patch. **(B)** Energy vs. Contrast per patch. **(C)** Homogeneity vs. Dissimilarity per patch. SVM, Support vector machine; RCM, reflectance confocal microscopy.

Fig. 11 Image processing steps for the identification of individual cells. A median filter and a local normalization were applied to the image with the ROI mask, followed by a Sato filter. Its output was filtered with a median filter and locally normalized, and a Gabor filter was applied to it. A threshold was applied on the output after histogram equalization and small blobs were removed with a connected components analysis. The result was then skeletonized, and spurious branches were removed. Image contrast was adjusted for easier visualization. Scale bar = 50 μm . ROI, region of interest.

Fig. 12 Post-processing steps. **(A)** The skeleton obtained after the previous step was cleaned, and contours were detected. Small contours were removed, as well as long contours close to the border with the background. The remaining contours were divided into two groups: small and big contours. Big contours were filtered again to improve the detection locally. The new resulting contours were then combined to the small contours and their centers were detected. In pink, some contours where two cells were merged are highlighted. **(B)** Example of large contours improvement for a *stratum granulosum* image. **(C)** Example of large contours improvement for a *stratum spinosum* image. Image contrast was adjusted for easier visualization. Scale bar = 50 μm .

Fig. 13 **(A)** Process of creation of synthetic RCM image of the *stratum spinosum* created using a hard-core process. First, points set apart with a minimum set distance from each other are created. Second, a tissue mask is created using Bezier curves and only the points within the mask are kept. Third, these points are used as seeds to initiate a Voronoi tessellation. Lastly, different noise levels are added to the image. **(B)** Side-by-side comparison of a real RCM image and a synthetic RCM image. RCM, Reflectance confocal microscopy.

Fig. 14 Detection accuracy evaluation steps. A marker-controlled watershed was applied to the detected cell centers and the resulting labels were compared to the manually detected ground truth, in pink. The returned metrics were precision and accuracy. Image contrast was adjusted for easier visualization. Scale bar = 50 μm .

Fig. 15 Diagram of the FIAP used for keratinocytes detection. The sections are color coded as follows: blue, ROI identification; pink, individual cells identification within the ROI; teal, post-processing steps. FIAP, Full image analysis pipeline; ROI: region of interest.

Fig. 16 The median cell area \pm standard error of mean per age group, epidermal layer and body site reflect the dynamic maturation and ageing of the epidermis. * Indicates that the median cell area is significantly different between the SG and the SS for a body site and age group. # Indicates that the median cell area is significantly different between the cheek and the volar forearm per age group and epidermal layer. SG, *stratum granulosum*; SS, *stratum spinosum*.

Fig. 17 Median cell area per participant on **(A)** the volar forearm (SG $R^2 = 0.465$; SS $R^2 = 0.00693$) and **(B)** the cheek (SG $R^2 = 0.0324$; SS $R^2 = 0.0708$) colored by age group. We fit a linear regression for each epidermal layer and body site but plot it only when significant. SG, *stratum granulosum*; SS, *stratum spinosum*.

Fig. 18 (A) Median SC thickness per participant (children $R^2 = 0.187$), **(B)** Median SPE thickness per participant (children $R^2 = 0.279$). A linear correlation with age was not significant in the adult group for either the SC or the SPE thickness. SC, *stratum corneum*; SPE, supra-papillary epidermis.

Fig. 19 Median cell aspect ratio per participant. SG, stratum granulosum; SS, *stratum spinosum*.

Fig. 20 Average probability distribution of the number of Delaunay nearest neighbors per cell for **(A)** SG on the volar forearm, **(B)** SS on the volar forearm, **(C)** SG on the cheek, and **(D)** SS on the cheek per age group. SG, *stratum granulosum*; SS, *stratum spinosum*.

Fig. 21 Probability distribution of the number of Delaunay nearest neighbors per cell for the cooperators model (in green), and defectors model (in red).

Fig. 22 Distribution of skewness per age group for **(A)** SG on the volar forearm, **(B)** SS on the volar forearm, **(C)** SG on the cheek, **(D)** SS on the cheek. In red, skewness of the probability distribution of the Delaunay nearest neighbors for the defectors model. In green, skewness of the probability distribution of the Delaunay nearest neighbors for the cooperators model. SG, *stratum granulosum*; SS, *stratum spinosum*.

Fig. 23 (A) Synthetic RCM images and its segmentation ground truth. **(B)** Tested U-net architectures segmentations on a synthetic RCM image. **(C)** Real RCM images and its segmentation ground truth. **(D)** Tested U-net architectures segmentations on a real RCM image. RCM, Reflectance confocal microscopy.

Fig. 24 DermoGAN architecture. The first task maps RCM images to the unpaired synthetic binary images. Whereas the second task learns the structure RCM images of the epidermis by translating Gabor filtered RCM images into binary images. RCM, Reflectance confocal microscopy.

Fig. 25 Comparison of the two cycle-GAN based approaches and the proposed DermoGAN. Manually obtained ground truth in the form of cell centers in green. sDermoGAN outperforms both methods.

Fig. 26 RCM image analyzed with the 5 presented algorithms. DermoGAN outperforms 4 out of 5 other methods on all images, and outperforms FIAP on 6/9 images. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

Fig. 27 DermoGAN can be extended using retraining to images acquired through different imaging techniques and outperforms traditional thresholding algorithms. **(A)** Input florescence microscopy images. **(B)** Output of DermoGAN applied to image A. **(C)** Application of 17 thresholding approaches to the same image.

Fig. 28 DermoGAN2 was trained entirely on synthetic images. **(A)** Synthetic non-confluent images created using the SIMCEP software. **(B)** Canny-filtered non-confluent images created using the SIMCEP software. **(C)** Binary non-confluent images.

Fig. 29 DermoGAN2 applied to an image of confluent BV-2 cells (on the left), resulted in accurate detection of cells (on the right). Manually determined cell centers were plotted on DermoGAN2 output in yellow.

Fig. 30 DermoGAN2 applied to an image of SK-BR-3 cells **(A)** resulted in accurate detection **(B)** of cells compared to thresholding methods **(C)**.

Fig. 31 DermoGAN2 applied to a mass spectroscopy image (on the left) resulted in accurate cell detection (on the right) with merging of adjacent cells.

Fig. 32 Histology image of healthy skin. Epidermis appears as a purple connected layer, while the dermis appears as a sparsely populated pink tissue.

Fig. 33 Application of DermoGAN3 on a histology image of healthy skin. **(A)** Original histology image. **(B)** Color normalized image. **(C)** Output of DermoGAN3 using the color normalized image. **(D)** Binarized output of DermoGAN3.

Fig. 34 Flowchart for the use of the developed graphical user interface including all developed methods.

Fig. S1 Gabor filters applied to the RCM image during the FIAP. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

Fig. S2 Synthetic images used in the training of the models. On the left, a synthetic binary image used in the DermoGAN training, and on the right synthetic RCM images of different noise levels and cell sizes used in U-net training. RCM, Reflectance confocal microscopy.

Fig. S3 Structure of the generator and discriminator networks used in the cycle-GAN and DermoGAN approaches.

Fig. S4 To obtain keratinocytes positions. We apply the $G_{A2B}: A \rightarrow B$ network to locally normalized RCM image and obtain an incomplete cell identification, which is then cleaned by closing any holes in the detected membrane and the outside contour, and finally the cell identification is refined using StarDist algorithm. RCM, reflectance confocal microscopy

List of Tables

Table 1 Epidermal layer classification algorithms in literature.

Table 2 Some skin lesions identification algorithms in literature.

Table 3 Study participants repartition per age group, gender, and Fitzpatrick phototypes.

Table 4 Cell detection accuracy on 6 *stratum granulosum* and 3 *stratum spinosum* RCM images for Expert 1. Ground truth in the form of cell centers is represented in green on each image. Median data shown as median (\pm 1 standard deviation). RCM, Reflectance confocal microscopy.

Table 5 Cell detection accuracy on 7 RCM images of the *stratum granulosum* for two different experts. Data shown as median (\pm 1 standard deviation). RCM, Reflectance confocal microscopy.

Table 6 Parameters for each step on the FIAP for RCM images of the SG and SS. FIAP, Full image analysis pipeline; RCM, Reflectance confocal microscopy. SG, *stratum granulosum*; SS, *stratum spinosum*.

Table 7 Average Fisher-Pearson coefficient of skewness of the probability distribution of Delaunay nearest neighbors per age group. SG, *stratum granulosum*; SS, *stratum spinosum*.

Table 8 Accuracy metrics for the tested U-net models.

Table 9 Comparison between the FIAP, Cell Cutter algorithm, and the U-net architecture. FIAP, Full image analysis pipeline.

Table 10 Comparison of median F1-score (computed with d-accuracy) for all six tested approaches.

Table 11 Comparison between the two main proposed methods for keratinocytes detection and manual cell identification.

Table 12 Summary of the outputs of each method.

Table S1 Cell detection accuracy on 2 *stratum granulosum* RCM images of dark- pigmented skin for Expert 1 using FIAP. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

Table S2 Comparison of all accuracy metrics for all six tested approaches (in %).

Table S3 Cell detection accuracy on 2 *stratum granulosum* RCM images of dark- pigmented skin for Expert 1 using DermoGAN. RCM, Reflectance confocal microscopy.

Table S4 RCM to histology translation results for a trained cycle-GAN and a trained DermoGAN. RCM, Reflectance confocal microscopy.

Contents

Analyse et caractérisation de la structure tissulaire de l'épiderme à partir d'imagerie confocale.....	3
Analysis and characterization of the tissue structure of the epidermis from confocal imaging	4
Acronyms.....	5
List of Figures.....	7
List of Tables	11
Contents	14
Chapter 1 Introduction: Reflectance confocal microscopy and its uses in skin research, available computational methods, and motivation for the work	17
1.1 Introduction to Reflectance confocal microscopy	17
1.2 RCM for skin research	20
1.2.1 Epidermal structure.....	20
1.2.2 Skin maturation and aging.....	26
1.2.3 Skin diseases.....	27
1.2.4 Cosmetology	28
1.2.5 Limitations of RCM use for the study of skin	28
1.3 Existing computational methods for RCM images analysis	29
1.3.1 Automatic identification of epidermal layers	29
1.3.2 Automatic identification of pigmented skin lesions	33
1.3.3 Automatic identification of cells	36
1.4 Research gaps and opportunities in skin RCM image analysis	37
Chapter 2 State of the art: cell detection in biomedical images.....	39
2.1 Traditional methods	39
2.1.1 Thresholding.....	39
2.1.2 Edge detection-based methods	40
2.1.3 Region-based cell segmentation methods.....	41
2.1.4 Partial differential equations-based methods.....	41
2.1.5 Watershed-based methods	42
2.1.6 Marked-point processes	43
2.1.7 Clustering-based methods.....	43

2.2	Deep learning-based approaches.....	45
2.2.1	CNNs	46
2.2.2	U-net	47
2.2.3	GANs and cycle-GANs	49
2.3	Multi-task learning.....	51
Chapter 3 A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features.....		
3.1	Full Image Analysis Pipeline (FIAP).....	56
3.1.1	Identification of the region of interest	56
3.1.2	Identification of individual cells.....	61
3.1.3	Post-processing	64
3.2	Data	67
3.3	Synthetic images	69
3.4	Accuracy evaluation.....	71
3.4.1	Accuracy on synthetic RCM images	73
3.4.2	Accuracy on real RCM images.....	74
3.5	Application of the FIAP for the study of age and body site-dependent changes in epidermal structure.....	81
3.5.1	Statistical analysis.....	82
3.5.2	Cell geometry evolution with age and epidermal layer.....	83
3.5.3	Cell topology evolution with age and epidermal layer.....	87
3.6	Comparison to machine learning based approaches	94
3.7	Discussion & Conclusion.....	101
Chapter 4 DermoGAN: An unsupervised multi-task deep-learning architecture for cell segmentation.....		
4.1	Methods.....	108
4.1.1	Identifying keratinocytes on RCM images with DermoGAN	108
	Generator and Discriminator architecture	109
	Multi-task approach.....	111
4.1.2	Refining the results with star-convex polygons detection.....	112
4.2	Experiments and Results.....	112
4.2.1	Dataset	112
4.2.2	DermoGAN implementation details	113
4.2.3	Comparison to other automated methods	113

4.2.4	Keratinocytes identification results	114
4.2.5	Extension to other image acquisition techniques	119
4.2.6	Retraining the model with only synthetic images	121
	DermoGAN2 on cell culture images	122
	DermoGAN2 on mass spectroscopy images	123
4.2.7	DermoGAN on histology images	124
	Dataset	125
	Results	126
4.3	Discussion & Conclusion.....	127
Chapter 5	Conclusion and Perspectives.....	130
5.1	Summary of the contributions.....	130
5.2	Implementation in a ready-to-use GUI for cell segmentation.....	133
5.3	Perspectives.....	135
	Bibliography	139
	Appendix	151
	Gabor filter bank	153
	Application of the FIAP to RCM images of dark-pigmented skin	154
	DermoGAN.....	155
	Accuracy metrics of the application of DermoGAN compared to those of 5 other approaches.....	157
	Application of DermoGAN to RCM images of dark-pigmented skin	158
	Thresholding approaches in ImageJ used in the evaluation of the accuracy of DermoGAN applied to fluorescence microscopy images.....	159
	Transforming RCM into histology and <i>vice versa</i>	161
	Publications	164
	Journals	164
	Conference proceedings.....	164
	Conference talks	164
	Patent Applications.....	166

Chapter 1 Introduction: Reflectance confocal microscopy and its uses in skin research, available computational methods, and motivation for the work

1.1 Introduction to Reflectance confocal microscopy

Biopsies followed by microscopic analysis of skin are the gold standard for the morphological investigation of skin¹. Unfortunately, biopsies are invasive, leading to local inflammation which may alter the original morphology, are not reproducible on the same body site, and can be traumatizing for the patient when done repeatedly. They may also raise ethical questions, such as for cosmetic testing² or the study of healthy infant skin. These limitations to biopsies in combination with advances in optics have allowed the development of non-invasive techniques for the imaging of skin, including optical coherence tomography (OCT)^{3–5}, magnetic resonance, and reflectance confocal microscopy (RCM)^{6,7}.

RCM allows for real-time *in vivo* visualization of the epidermis and the upper parts of the dermis at cellular level^{6,8}, providing information on the morphology and topology of the skin⁹. It enables the repeated sampling of the tissue without damage to the observed area, making it a technique of choice for the observation of the dynamic changes of the upper parts of the skin over-time and for the quantitative and qualitative study of the cellular structures involved in the makeup of the skin barrier^{10–12}. Unfortunately, RCM is limited by the maximum observable depth in the tissue before the signal-to-noise ratio becomes too low to acquire any significant information, but it provides information faster than microscopic analysis of a biopsy sample.

The confocal microscope was invented by Marvin Minsky in 1957¹³. Two reflectance confocal microscope versions are currently commercially available, a handheld *in vivo* skin imaging

microscope¹⁴ and a wide-probe RCM¹⁵. They offer a horizontal resolution of $0.5 - 1 \mu\text{m}$ and a vertical resolution (optical section thickness) of $3 - 5 \mu\text{m}$, to a depth of $150 - 200 \mu\text{m}$ depending on the observed site¹⁵.

RCM relies on the detection of signals arising from the reflection of light at the interface between microstructures with different indices of refraction. In skin, such microstructures are cell membranes, collagen and keratin fibers, melanosomes, and intracellular organelles¹⁶. The closer the size of the organelle is to the wavelength of the light source and/or the higher its refractive index is compared to its surroundings, the brighter it appears⁸ on an RCM image.

RCM is a suitable technology for examining structures *in vivo* since the energy from the incident light is sufficient to generate a signal but not so strong as to trigger a photobiological process. Therefore, RCM enables visualization of living cells without causing a disruption or an alteration to their structure or function.

RCM stacks are gray-scale images acquired at sequential depths starting from the skin surface. Their orientation is perpendicular to the vertical sections typical in histopathology.

An *in vivo* reflectance confocal microscope is composed of a light source, apertures, lenses, a detector, and a beam-scanning mechanism¹⁷ (see Fig. 1). The light source is typically a near-infrared laser. In the reflectance setup of a confocal microscope, the objective lens plays a dual role of focusing the illumination onto a specific spot of the sample and collecting the reflected signal. A detection pinhole exclusively allows light reflected from the illuminated point on the sample to reach the detector while effectively rejecting scattered light. By systematically scanning the illumination spot over the sample, a complete image is formed through the reconstruction of the acquired data.

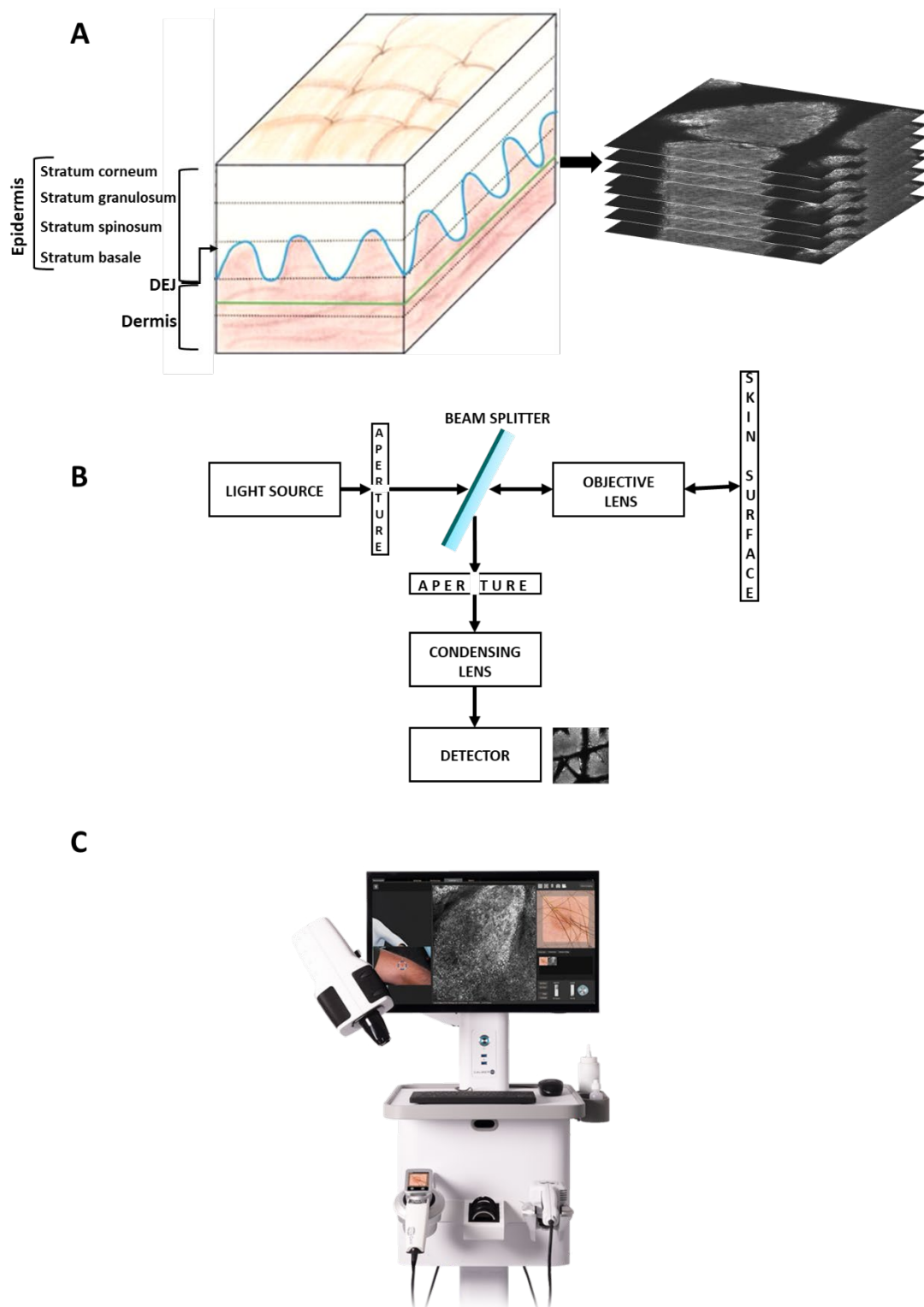


Fig. 1 (A) RCM stacks are gray-scale images, orthogonal to the vertical sections typical of skin histology. The left panel shows an illustration of the undulating DEJ in blue. The depth limit of RCM is delineated in green, and RCM optical sections are represented with black dotted lines. The right panel shows a stack of images at sequential depths corresponding to these optical sections. **(B)** Diagram of an *in vivo* reflectance confocal microscope. **(C)** Image of a VivaScope 1500 (Lucid, Inc., Rochester, New York) reflectance confocal microscope.

RCM, Reflectance confocal microscopy; DEJ, Dermal epidermal junction.

1.2 RCM for skin research

RCM has become a valuable tool in dermatology to visualize skin structure in a wide range of applications, *e.g.*, study of healthy skin maturation and ageing, diagnosis and monitoring of skin inflammatory diseases and cancer, delimitation of lesions, and evaluation of treatment efficacy.

1.2.1 Epidermal structure

The epidermis is an avascular keratinized stratified squamous epithelium generally made of four distinct layers. From superficial to deepest, these layers are called *stratum corneum* (SC), *stratum granulosum* (SG), *stratum spinosum* (SS), and *stratum basale* (SB). In the soles and palms, a thicker epidermis is observed, with an additional fifth layer between the cornified and the granular layers called the *stratum lucidum*.

Most cells in all layers below the SC are referred to as keratinocytes, thus named due to their involvement in the manufacturing and storing of keratin intermediate filaments. In contrast to the viable keratinocytes, the SC is made of dead but enzymatically active cells called corneocytes¹⁸. Throughout the lifetime of a person, these cells are shed, and replaced by others from the lower layers. The process starts in the basal layer, where cells are continuously produced (by stem cells and transient amplifying cells), lose their attachment to the basal membrane, and migrate towards the upper layers, while undergoing differentiation towards final cell death in a process called cornification.

RCM can be used to observe the epidermal layers, the dermal-epidermal junction (DEJ) and the upper layers of the dermis¹¹, thus allowing the computation of several quantitative descriptors of skin structure, such as keratinocyte density, number of basal keratinocytes around each dermal papilla, length of DEJ, and circumference of dermal papillae which are small, finger-

like projections of the dermis into the epidermis. They serve multiple functions, including increasing the contact surface between the dermis and the epidermis, supplying nutrients, facilitating sensation, supporting hair and nail growth, contributing to temperature regulation, and forming unique fingerprint patterns. Measuring these parameters on RCM images enables the quantitative study of skin structures and their evolution over time, for example as a response to different stimuli. Besides the geometrical parameters we can also extract information about the topological organization of the epithelium, for example the distribution of the number of nearest neighbors to each cell, an important factor in determining molecular exchange rates between neighboring cells⁹.

The top slices of RCM stacks represent the SC, which appears as large bright areas forming islands surrounded by dark empty areas (Fig. 2A). These dark areas are due to grooves called skin micro-relief lines⁶, while the bright signal in the island structure is due to the high reflectance of keratin. The cells are anucleated dead corneocytes, made primarily of aggregated keratin filaments embedded in a lipid matrix¹⁸, polygonal in shape, and 10 – 30 μm in size⁸.

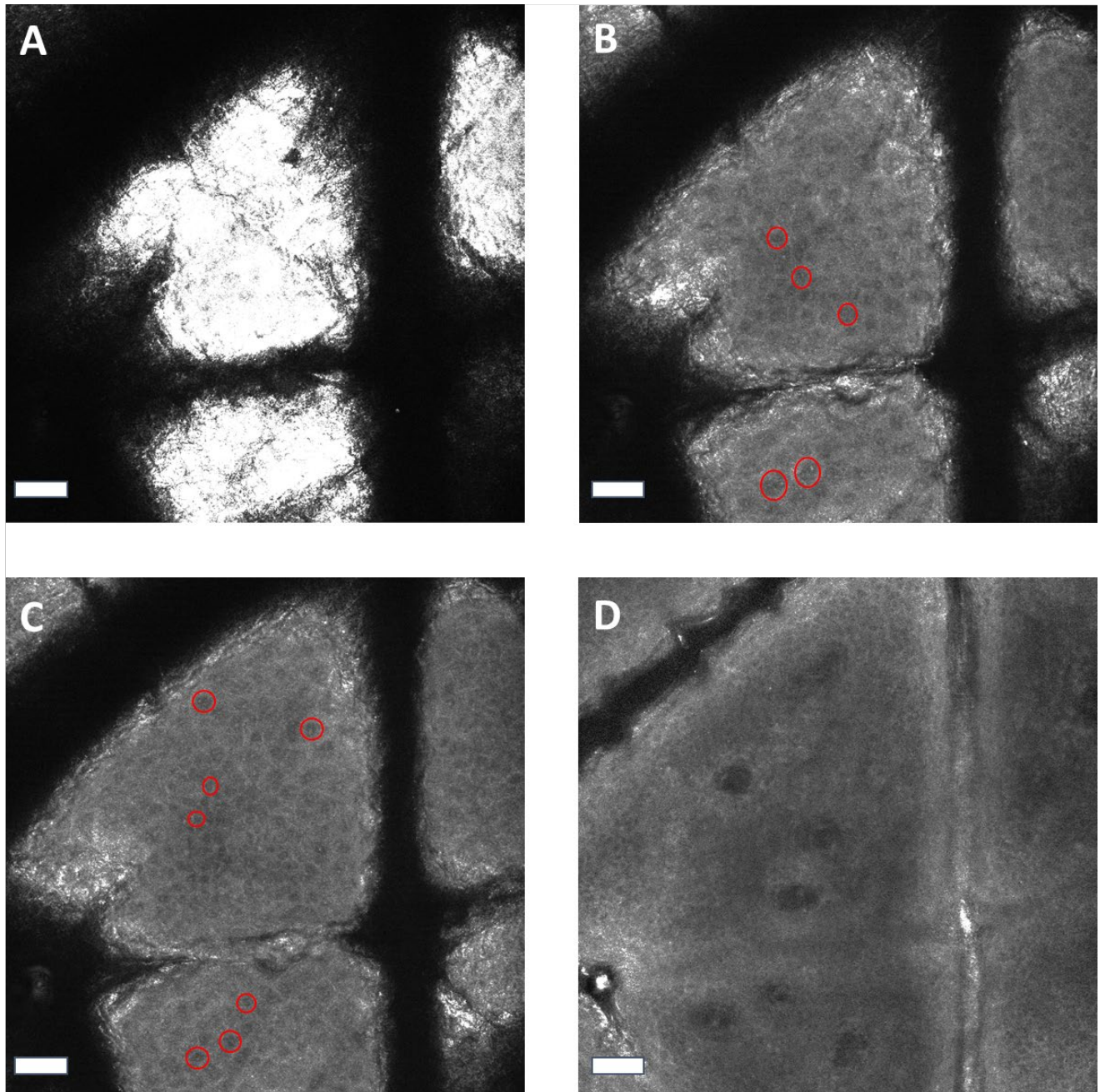


Fig. 2 Representative RCM images of minimally pigmented skin acquired at depths corresponding to the (A) *stratum corneum*, (B) *stratum granulosum*, (C) *stratum spinosum* and (D) *stratum basale*. In red, examples of granular and spinous keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy. Contrast was adjusted for clearer images.

SC thickness is an important factor involved in skin barrier function^{12,18}. The thicker the SC, the more difficult it is for a noxious substance to penetrate the viable parts of the epidermis (or equivalently for water to transverse the epidermis and evaporate, potentially leading to tissue desiccation). It is 12 to 208 μm thick depending on the body site¹⁹. Moreover, corneocytes

provide a mechanical strength to the skin surface and are involved in protecting the lower layers against ultraviolet (UV) radiation, while the lipid matrix is important in maintaining skin permeability^{20,21}. SC thickness can be calculated by the depth difference of the uppermost and the lowest optical sections that contain SC structures.

The *stratum granulosum* (Fig. 2B) and *stratum spinosum* (Fig. 2C) are respectively the second and third layer in the epidermis from the skin surface. They are composed of keratinocytes arranged in a honeycomb pattern in minimally pigmented skin and a cobblestone pattern in heavily pigmented skin¹¹ (Fig. 3B, 3C). In minimally pigmented skin, the cells are characterized by a dark center, grainy cytoplasm due to organelles and microstructures¹⁰, and surrounded by bright membranes^{8,11}. In heavily pigmented skin types, due to the high melanin-content in melanosomes, which gives a strong reflectance signal, we observe bright keratinocytes separated by a dark contour²². Viable keratinocytes are found at depths of 20 – 100 μm and are about 10 – 15 μm in size⁸. Cells are typically larger in the granular layer than in the spinous layer⁹ where they have a higher density. Indeed, as the keratinocytes further differentiate while climbing towards the surface, they get wider and flatter.

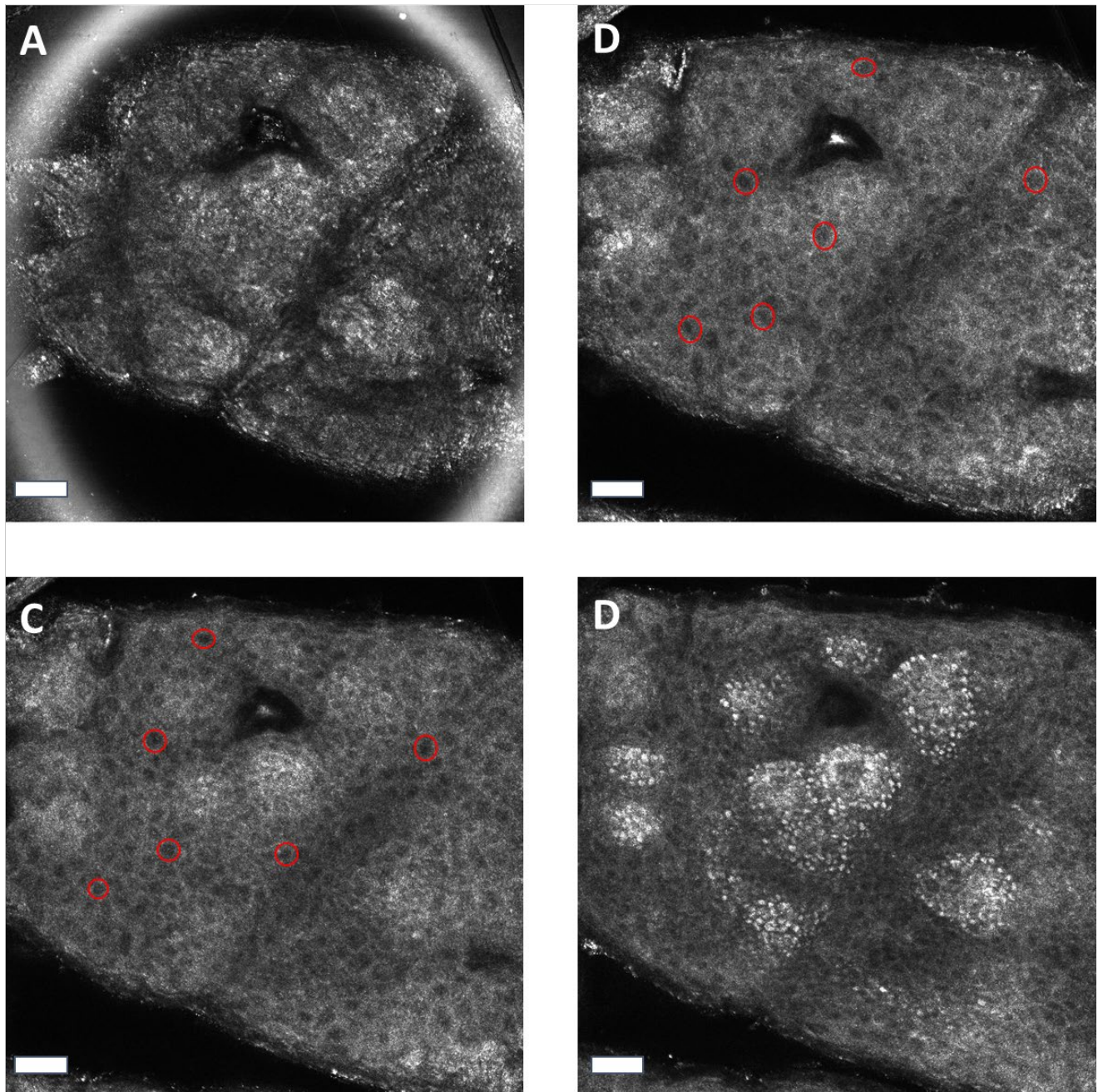


Fig. 3 Representative RCM images of a heavily pigmented skin acquired at depths corresponding to the (A) *stratum corneum*, (B) *stratum granulosum*, (C) *stratum spinosum* and (D) *stratum basale*. In red, examples of granular and spinous keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy.

Towards the basal layer of the epidermis, the cells appear similar in shape but smaller in size compared to the two previous layers (Fig. 2D). In contrast to the other layers, the basal keratinocytes make a monolayer. These cells are precursors of the keratinocytes in the upper

layers and appear brighter than them due to the presence of melanin, which has a high reflectance^{7,23}. Melanin is made by melanocytes scattered through the basal layer and then transferred to the keratinocytes²⁴. The cells of this layer are adherent to a collagenous membrane that separates the epidermis from the dermis called the basement membrane.

The thickness of the viable epidermis can be calculated as the depth difference between the optical sections at which we observe discernable viable keratinocytes in the *stratum granulosum* and that at which the top of the DEJ appears in the *stratum basale*.

The undulating DEJ (Fig. 1A) separates the epidermis from the dermis and is located at 50 – 100 μm below the skin surface.

Sometimes, bright areas can be observed on RCM images at various layers. They may arise from the keratin in hair shafts (Fig. 4A) or from clustered keratinocytes called mottled pigmentation² (Fig. 4B).

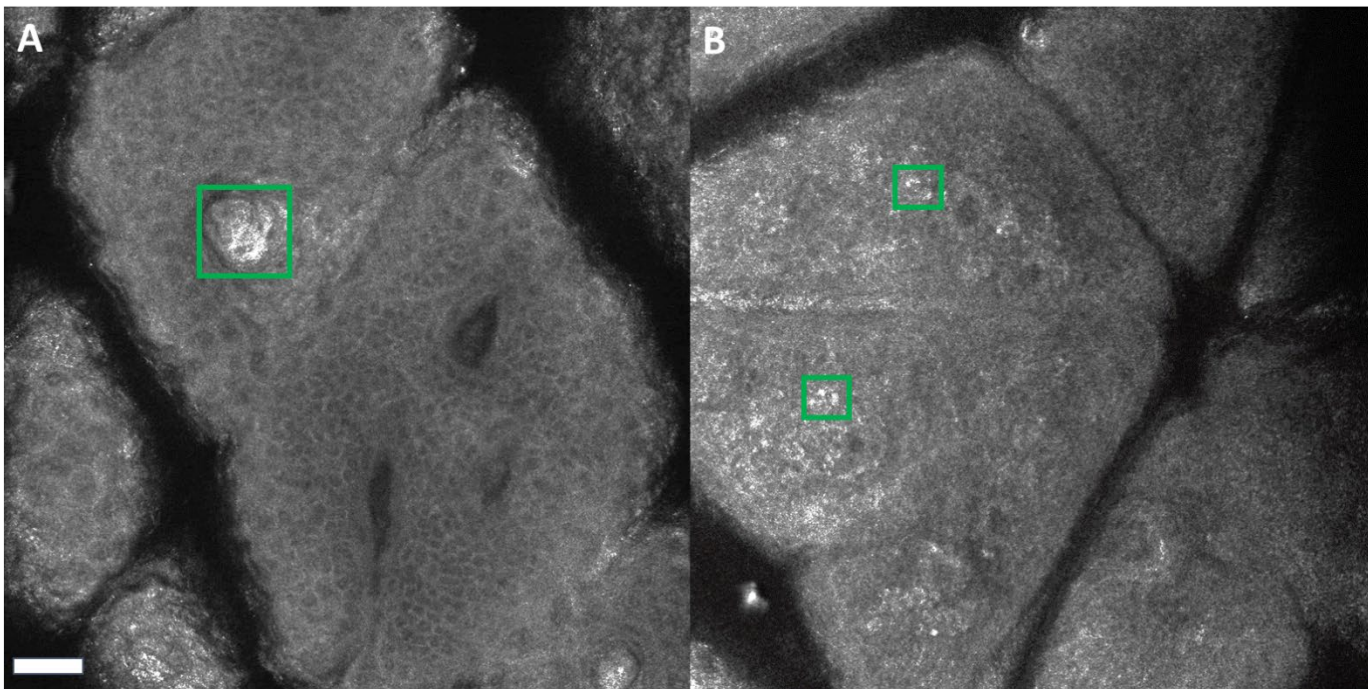


Fig. 4 RCM images with bright spots due to (A) keratin in hair shafts and (B) clustered keratinocytes. Scale bar = 50 μm . RCM, Reflectance confocal microscopy.

1.2.2 Skin maturation and aging

Skin structure matures and ages, changing throughout our lifetime^{25,26}. These changes are observable on RCM images.

Infant and adult skins are structurally different, from skin surface appearance to epidermal layers thicknesses, dermal extra-cellular matrix structures, and skin composition²⁷. Indeed, infant skin has thinner more abundant micro-relief lines^{9,12} which plays a role in the absorption of topically applied products, as micro-relief lines may act as reservoirs for the products, therefore affecting their permeability kinetics. Additionally, the SC is about 30% thinner in children than in adults, and the supra-papillary epidermis is 20% thinner^{12,28}.

The structural differences between adult and children skin translate into functional differences, *e.g.*, trans-epidermal water loss (TEWL) which reflects the quality of the skin barrier function, is significantly higher in children and decreases during childhood until reaching adult values¹². Moreover, children epidermal cells turnover is faster²⁸, leading to smaller keratinocytes and corneocytes^{9,27}, and therefore denser epidermal layers. With age, cell perimeter and surface area increase, as well as overall epidermal thickness and individual layer thickness.

Skin continuously changes and keeps maturing during adulthood. RCM can be used to document and quantify subclinical microscopic changes in skin^{11,22,29–31} and their correlation to early manifestations of skin ageing²² and the effects of cosmetic products on the reduction of these visual markers of age, *e.g.*, wrinkles, thinning of the skin, hyper-pigmentation spots, and loss of elasticity. Indeed, with age, keratinocytes honeycomb pattern becomes more irregular, and mottled pigmentation becomes more frequent^{2,30}. Collagen fibers also become more compact under the DEJ. A decrease in the number of dermal papillae in aged skin has also been observed on RCM images^{30,32}.

These changes in skin structure and morphology are more frequent and obvious in sun-exposed areas where photoaging, caused by exposure to UV radiation, accelerates the natural biochemical and physiological ageing processes. Indeed, exposure to solar UV radiation is a well-documented risk factor of overall skin damage and skin cancer, and an accelerant of skin aging. RCM allows the longitudinal assessment of the effects of sun exposure on skin³³ as it can be performed repeatedly and non-invasively on the same areas over time. RCM images have shown that total epidermal thickness and keratinocyte density are greater in sun-exposed areas and on the face¹⁰ and their honeycomb pattern disturbed³³, two microscopic signs of early skin ageing.

1.2.3 Skin diseases

Descriptive features of skin inflammatory diseases can be observed *in vivo* using RCM, *e.g.*, psoriasis³⁴⁻³⁶ and allergic contact dermatitis^{20,37,38}, for both diagnosis and disease progression monitoring. As it is faster than performing a biopsy followed by microscopic analysis of a sample, RCM has potential to be integrated as an initial step in the clinical diagnosis.

Psoriasis is characterized by a thickening of the SC and viable epidermis; both features quantifiable by RCM. RCM has been used in patients with psoriasis to document thinning of the granular layer⁸, increase in the number and size of dermal papillae, and increase in keratinocytes size and brightness.

Diagnosis of allergic contact dermatitis can be guided by RCM. Some of its characteristics observable in RCM are disrupted SC, vasodilatation, increased epidermal thickness and detached corneocytes^{20,38}.

RCM is also useful in the diagnosis of cancer, as it can show, among other things, DEJ disarray and atypical cells, which have been identified as 2 of 18 RCM features useful in the diagnosis of malignant melanoma³⁹. RCM has also been shown to be useful in the diagnosis of non-

melanocytic lesions, using features such as parakeratosis, irregular honeycomb pattern in the SS and SG, and spindle shaped cells with dendritic branches infiltrating the epidermis⁴⁰. In combination with dermoscopy, RCM has been shown to increase diagnosis specificity and reduce the number of unnecessary excisions of potential melanocytic lesions^{41–43} by 2-3 times⁴⁴, which not only has a positive impact on patients, but also on the healthcare system and costs associated with the management of skin cancer.

In addition to diagnosis, RCM can be used in the examination of tumor/disease spread. Indeed, its large field of view, much larger than that of a biopsy, can be used to determine lesion margins, which may be useful when monitoring the spreading of a disease and/or its shrinkage post non-surgical treatment as a measure of its efficacy, or to guide surgical excision of a cancerous lesion.

1.2.4 Cosmetology

RCM is useful in the quantification of the impact of cosmetic formulations on skin, *e.g.*, retinoic acid, retinol, cleansers, and moisturizers^{45–50}. This is particularly useful, as biopsies cannot be ethically performed for these tasks. RCM, therefore, offers a non-invasive alternative to link the effects of cosmetic products on epidermal structure and skin surface appearance.

1.2.5 Limitations of RCM use for the study of skin

RCM main limitation is its maximum penetration depth, which varies by body site and tissue type, before the signal-to-noise ratio is too low to extract any useful information. In skin, we can only observe the epidermis, DEJ, and upper layers of the dermis. Indeed, while refractive microstructures are abundant in the dermis, light intensity and coherence drop exponentially with depth.

RCM is also limited, compared to histology, by the lack of automated methods for features extraction. Indeed, analysis of RCM images is mainly performed manually, which is time-consuming, tedious, and subject to inter-and intra- expert differences. Additionally, it requires 4-6 months of training¹⁵. Developing standardized validated methods for the automated analysis of RCM images could help it become more wide-spread⁵¹, both in skin research and clinical practice.

Using RCM may be inadequate when multiple invasive tests have to be performed. Indeed, one biopsy sample can be used to perform multiple analysis, *e.g.*, DNA sequencing, cytogenetic testing, on top of its microscopic examination which is not possible when using RCM.

1.3 Existing computational methods for RCM images analysis

Some methods have been developed for the automated extraction of certain descriptors from RCM images, *e.g.*, identification of the epidermal layers, of pigmented skin lesions, and of cells. However, to date none of these methods have been integrated in clinical care.

1.3.1 Automatic identification of epidermal layers

Multiple attempts at automating epidermal layer identification have been made using machine learning^{52–56}. The maximum accuracy obtained by these algorithms is reported to be 88%.

Somoza *et al.*⁵⁵ achieved a classification accuracy of 54% for epidermal layer identification using an unsupervised texton-based method (see Fig. 5). Textons are minimal building blocks of vision with no strict definition in literature. The approach involved creating a texton library of micro-structures by convolving RCM images with 10 Leung-Malik filters that matched the size of a keratinocyte. A principal components analysis (PCA) was then applied to reduce the dimensionality of the texton-space to 3, followed by K-means clustering with 15 clusters. This

texton library was applied to RCM images, and the results were projected onto the 3 selected PCA axes. Based on the Euclidean distance, each pixel on the RCM images was classified as 1 of the 15 textons and represented as a 15-dimensional texton histogram. The histogram dimension was then reduced to three using a second PCA. Finally, the classification of each pixel was obtained by applying a K-means clustering with five clusters, four of them representing the four layers of the epidermis and the fifth representing the dermis.

The texton-based approach has the potential to be adapted to other classification tasks, such as evaluating the effectiveness of treatments for different skin diseases or studying skin aging and maturation, by expanding the texton library to include more features. However, this approach can also be improved by incorporating higher-level information and features. Currently, this method does not consider cellular characteristics or the presence of reflective or darker surfaces, which are factors that experts take into account when manually identifying epidermal layers in RCM images. Additionally, this was a pilot study, conducted on only three stacks and assumed that each image contained only one epidermal layer, which is often not the case.

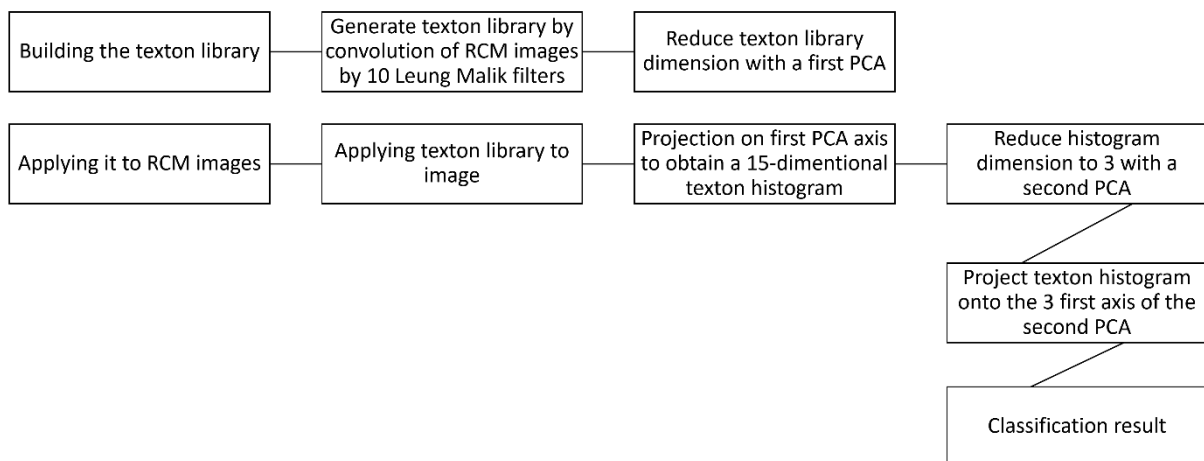


Fig. 5 Diagram of epidermal layer classification used by Somoza *et al*⁵⁵.

Hames *et al.*⁵⁴ used a bag of features approach to classify RCM images into four categories: SC, viable epidermis, DEJ, and papillary dermis. Four features were identified based on prior

knowledge of RCM images: 1) a visible honeycomb pattern of viable keratinocytes indicating viable epidermis, 2) the presence of bands characteristic of basal cell/dermal papillae indicating the DEJ, 3) the absence of *stratum basale* features, and 4) visible papillae indicative of the papillary dermis. They created a feature dictionary from small image patches and used it to represent each test image as a histogram of counts of visible features. An L1 regularized logistic regression was used to classify each histogram into 1 of the 4 categories. The authors achieved a classification accuracy ranging from 62.9% to 95.6%, depending on the epidermal layer, body site, and phenotype. However, not all phenotypes or body sites were included in the study, and diseased skin was not covered. The method assumed a single epidermal layer per image, which is often not the case.

Kaur *et al.*⁵³ proposed a hybrid deep learning approach to classify RCM images into five categories: the four epidermal layers and the dermis. Initially, each RCM image underwent convolution using a 48-filter bank. Then, a pre-built texton library was employed to represent each pixel with a labeled patch centered around it. Each pixel was linked to its eight closest neighbors, resulting in eight texton maps per image. The maps were then pooled by weight, and the resulting histograms were used to train a convolutional neural network (CNN) with empirically optimized parameters. The method achieved an accuracy of 82%, but was only tested on three RCM image stacks, and the results are limited by the features in the texton library. While building the library using a multi-resolution, multi-orientation filter bank increases the number of features compared to manual determination, it complicates the interpretation of each feature.

Bozkurt *et al.*^{51,52} proposed an automated method for classifying RCM images into epidermal layers based on a recurrent CNN, which allows the network to process related temporal or spatial data by feeding results back into the network. The authors introduced a Toeplitz structure to aid model interpretation by indicating which image the model decision was based on. The

model is an encoder-decoder that uses bi-directional recurrent units and Inception V3 networks⁵⁷. This approach achieved an accuracy of 88% in classifying RCM images into the epidermis, DEJ, and dermis. It was tested on a larger dataset than the previously mentioned methods for automatic identification of epidermal layers and incorporates higher-level information by considering three surrounding images when making a prediction. However, using a CNN leads to a loss of feature interpretability.

Table 1 summarizes the methods described for automated epidermal layer classification. Neural network-based approaches were overall more successful in correctly classifying images to epidermal layers than algorithms based on texture analysis. Direct comparison between the methods is not possible as they were not trained nor tested on the same images nor on the same number of images or types of observed tissue (healthy vs. lesional), although all used images were captured using a VivaScope 1500 (Lucid, Inc., Rochester, New York) reflectance confocal microscope (see Fig. 1C). Additionally, the accuracy metrics were calculated against a manually obtained image-level ground truth by different experts for each method. Only one study indicated using data including Fitzpatrick skin phototypes I to IV. Fitzpatrick phototypes form a system used to describe a person's skin based on its response to ultra-violet exposure on a scale from I to VI.

Table 1 Epidermal layer classification algorithms in literature.

Reference	Year	Method	Images database	Accuracy	Sensitivity	Specificity
55	2014	Unsupervised texton-based approach	Testing: 3 stacks of adult males Type of sample: Normal Site: volar forearm	54%	0.53	0.9
54	2016	Bag of features and logistic regression	304 stacks (54 volunteers, age 20-30 and 50-70) Type of sample: Normal Phototypes: I - IV Site: Dorsal and volar forearm	86%	0.84	0.92
53	2016	Hybrid deep learning	15 stacks (1500 images) Training: 12stacks Testing: 3 stacks Type of sample: Normal Site: -	82%	0.72	0.96
52	2017	Recurrent convolutional neural network	Training: 245 stacks Testing: 61 stacks Validation: 198 stacks Type of sample: normal, benign melanocytic, and diseased Site: arms, legs, torso	88%	0.87	0.94

1.3.2 Automatic identification of pigmented skin lesions

Multiple types of lesions are identifiable on RCM images, and different attempts at automatically identifying them have been made. We can distinguish two types of applications of these algorithms: 1) finding melanoma patterns and 2) distinguishing non-melanocytic lesions from melanoma.

One of the proposed algorithms to find melanoma patterns and identify melanocytic lesions on RCM images is based on a wavelet transform⁵⁸. It obtained moderate success with 55% of the melanomas and 47% of the benign nevi being correctly identified.

Another approach⁵⁹ aimed to determine the presence of melanoma by identifying patterns in the DEJ mosaics and classifying them into melanoma or non-melanoma with a sensitivity of 55 – 81% and a specificity of 81 – 89%, thus mimicking a clinician's understanding and reading of an RCM image. A more recent approach by Bozkurt *et al.*⁶⁰ used a multiresolution CNN to identify similar patterns as Kose *et al.*. It achieved 95% average specificity and 77% average sensitivity.

For the second application of distinguishing between melanocytic and non-melanocytic lesions, Halimi *et al.*⁶¹ proposed a Bayesian model to quantify RCM images reflectivity and classify images in two categories, healthy and lentigo patients, based on their reflectivity distribution. They obtained an accuracy of 98%.

Zorgui *et al.*⁶² obtained similar results, with an accuracy of 98% with a CNN. The CNN was trained on normalized resized RCM images with a pretrained Inception V3 model. Transfer learning was then used to apply the model to skin RCM images.

Finally, Bozkurt *et al.*⁶⁰ proposed a CNN inspired from the U-net architecture to identify six classes: non-lesion, artifact, meshwork pattern, ring pattern, nested pattern, and aspecific/patternless. This model was built on a dataset containing RCM images of both lesional and non-lesional skin. This method slides through RCM images with a sliding window with 75% overlap and applies three consecutive nested U-nets. This generates segmentations at different resolution levels. Each U-net model generates a probability map. The deepest U-net model only takes a sliding window as input, while the others use a concatenation of the up-sampled probability map at the higher level and sliding window. This model achieved 73% classification accuracy.

Introduction: Reflectance confocal microscopy and its uses in skin research, available computational methods, and motivation for the work

Table 2 summarizes some of these algorithms and the lesion they seek to identify. We do not aim to compare the different methods as they were built to identify different lesions, but to show the variety of identifiable lesions through the use and automated analysis of RCM images.

Table 2 Some skin lesions identification algorithms in literature.

Reference	Year	Lesion	Method	Image database	Ground truth type	Result
58	2011	Melanocytic lesions	Wavelet transform with Support Vector Machine	Training: 6147 images Testing: - Validation: 10122 images Type of sample: melanocytic skin lesion Site: -	Image-level classification	55% accuracy for melanomas detection 47% accuracy for benign melanocytic nevi detection
63	2010	Superficial spreading melanoma versus nevi	Pattern recognition algorithm	200 sections total Type of sample: 5 superficial spreading melanoma and 5 nevi Site: regions containing visually recognizable cells	Object-level detection (detection of pagetoid melanocytes)	100% accuracy is small pilot study
59	2016	Mosaics of the DEJ	SVM	Training: - Testing: - Validation: 20 RCM mosaics Type of sample: - Site: -	Object-level classification (meshwork, ring, clod, aspecific and background patterns)	55 – 81% sensitivity, 81 – 89% specificity
60	2018	Melanomas	Convolutional neural network	Training: 46 RCM mosaics Testing: 10 RCM mosaics Validation: - Type of sample:	Object-level classification (mesh, ring, nest, aspecific, artifact,	77% average sensitivity,

Introduction: Reflectance confocal microscopy and its uses in skin research, available computational methods, and motivation for the work

Reference	Year	Lesion	Method	Image database	Ground truth type	Result
				Melanoma suspicious patients Site: -	background patterns)	95% average specificity
61	2017	Lentigo	Bayesian model	Training: - Testing: - Validation: - Type of sample: 27 women with lentigo and 18 without Site: back of the hand	Image-level classification	98%, accuracy, 96% sensitivity, 100% specificity
62	2020	Lentigo	Convolutional neural network	Training: 312 images Testing: 56 images Validation: 60 images Type of sample: healthy and lentigo Site: -	Image-level classification	98% accuracy, 96% sensitivity, 100% specificity
60	2018	Morphological pattern	Nested U-net	Training: 46 RCM mosaics Testing: 10 RCM mosaics Validation: - Type of sample: Melanoma suspicious patients Site: -	Object-level classification (mesh, ring, nest, aspecific, artifact, background patterns)	73% accuracy

1.3.3 Automatic identification of cells

Individual cell location visible in RCM images provide important information in the assessment of skin health, but their manual identification is tedious, time-consuming, and subject to expert interpretation. To our knowledge, only one attempt has been made to automatically identify individual cells or nuclei in skin RCM images.

Gareau⁶⁴ attempted to automatically identify keratinocytes on RCM images. An error function reflectance profile was trained on labeled RCM images and then tested on other images to identify keratinocytes coordinates, with one set of parameters for both SG and SS images. All images belonged to the same stack which limits the validity of this method. The obtained keratinocyte density matched prior knowledge based on manual counts and was statistically validated, basing its accuracy on obtained cell density. The model supposes that the keratinocytes center is darker than the rim. This assumption fails on basal cells due to bright light-scattering melanosome caps over the nuclei. The method may be improved by training two separate models for the granular and spinous layers as their cells differ in size. It is also unclear how it behaves when tested on RCM images of other people of different ages, as keratinocytes size change with age.

In all models described above across all presented computational methods, results differed between minimally and heavily pigmented skin or were not tested in both cases.

1.4 Research gaps and opportunities in skin RCM image analysis

RCM offers the possibility to non-invasively quantitatively and qualitatively study healthy and diseased skin structure and topology. RCM can be, and has been, used in skin research and has the potential to be integrated in clinical practice. Unfortunately, the use of RCM is hindered by image quality, high noise, low contrast, and drop in signal-to-noise ratio, and by the need to manually analyze the images, which is tedious, time-consuming, and subject to human error. While attempts at automating descriptor extraction have been made, only one method⁶⁴ has been proposed for individual cell detection, an important first step in the quantitative analysis of any tissue, and a ready-to-use gold standard is not yet available. Therefore, research should

Introduction: Reflectance confocal microscopy and its uses in skin research, available computational methods, and motivation for the work

focus on methods allowing an easier translation of images into relevant quantifiable parameters, and on making the use of RCM easier, faster, and more accessible.

To address these issues, in this thesis: (1) We propose an automated full image analysis pipeline (FIAP) for the identification of cells on RCM images of the *stratum granulosum* and *stratum spinosum*. (2) We use the proposed method on over 5000 images of clinical study participants, ages 0 to 80 years old, to calculate parameters related to tissue architecture and relevant to skin physiology to validate the dynamic changes of skin maturation during childhood and skin aging in adulthood, examining the effects of age on skin in different body sites and epidermal layers. To our knowledge this is the first large-scale analysis of RCM images. (3) We contrast the FIAP, to a novel unsupervised deep learning architecture based on multitask cycle-generative adversarial networks (GANs), and show how this method, which does not require manual labeling for training, can be used, with or without retraining, and adapted on images acquired by different imaging techniques on multiple tissue types. (4) We finally explore methods for the translation of RCM images into histology images, and *vice versa*. We believe this could be of use to dermatologists and pathologists, more likely to be trained on the diagnosis of diseases using histology slides than RCM images, allowing them to unlock insights powered by confocal images through the lens of histology.

All the proposed methods were then integrated as tools in a ready-to-use user interface to facilitate their application for RCM segmentation and interpretation.

Chapter 2 State of the art: cell detection in biomedical images

Cell segmentation, *i.e.*, partitioning of a biomedical image into individual disjoint instances of cells, plays a crucial role in the understanding of healthy tissue, and the diagnosing and treating of diseases by providing information relating to cell morphology and physiology⁶⁵. However, it is a subjective and time-consuming task that could benefit from the use of computer-aided systems. This would allow for large-scale and systematic analysis of biomedical images, which would facilitate the quick evaluation of changes in cell features, *e.g.*, cell count, type, shape, and texture⁶⁶, over time and in response to different conditions.

Many methods have been developed for automated cell segmentation on biomedical images, dating back to the 1960s. All existing methods can be roughly grouped into three categories: (1) traditional methods, (2) deep learning-based approaches, and (3) multi-task learning.

2.1 Traditional methods

Traditional methods for cell segmentation make use of image processing methods coupled with optimization algorithms based on mathematical operations. These methods are well established, explicable, and interpretable, and can be divided into seven subcategories: thresholding methods, edge detection methods, region-based methods, partial differential equations (PDEs)-based methods, watershed-based methods, marked-point processes, and clustering-based methods.

2.1.1 Thresholding

Thresholding is the simplest most intuitive traditional segmentation method, particularly useful for images with high contrast between the background and objects of interest. The threshold

value which will be used to divide the image into two or more groups, can be set locally (adaptive thresholding⁶⁷) or globally, manually or automatically.

One of the most referenced thresholding method is Otsu's thresholding⁶⁸, where an automatic optimal threshold is set to minimize the intraclass variance among the region-of-interest, and the intraclass variance among the background. While this method has been used in cell segmentation, it is often not enough to obtain satisfactory results given the complexity of biomedical images^{69,70}, and so it is often integrated in a more complex segmentation pipeline⁷¹⁻⁷⁵. Indeed, thresholding requires there is a difference in intensity between background and foreground (cells), a condition often unsatisfied in biomedical images, making the satisfactory thresholding of the image complicated.

2.1.2 Edge detection-based methods

Edge detection-based methods focus on finding objects boundaries in images by determining the edge pixels located between the cells and the background, generally corresponding to an abrupt change or discontinuity in colors. Most edge detection methods use derivatives to detect object boundaries. Classical methods, like the Sobel-Feldman edge detector⁷⁶ and the Scharr filter⁷⁷, employ the first derivative which may be subject to noise and therefore inaccuracies, but are simple and allow for edge orientation detection.

Another one of these methods is the Canny edge detector⁷⁸ based on a gaussian smoothing filter to counter the effect of noise, followed by gradient intensity calculation and a hysteresis thresholding. This method improves the signal-to-noise ratio. Another solution to reduce the impact of noise on the detected edges is to use a filter, such as the Laplacian of Gaussian algorithm⁷⁹, also known as the Mexican Hat filter. It first smooths the image before calculating the Laplacian to detect object boundaries by looking for its zero crossings, this allows to test a

wider area around the boundary but causes artifacts at corners and curves. The use of the Laplacian also forbids the detection of edge orientation.

All of these methods have been applied for cell segmentation with various levels of success and their accuracy depends on the application.

2.1.3 Region-based cell segmentation methods

Region-based cell segmentation methods look for similarities, *e.g.*, intensity, rate of change in intensity, color, and texture⁸⁰, between adjacent pixels in the image and grouping them under a single label. These methods can embed region growing^{81–83} and/or splitting, but they all exploit the spatial context of the image. The first type (bottom-up approach) recursively grows an area by including similar neighboring pixels, while the second (top-down approach) starts by considering the entire image as one segment that is recursively split based on a set of criteria. Region-based segmentation methods are often combined with thresholding approaches in order to split overlapping cells^{74,75,81,84–86}.

Region and edge detection based methods are often hindered by the lack of sharp boundaries in biomedical images⁸⁷.

2.1.4 Partial differential equations-based methods

Image segmentation using PDEs is often based on the curve propagation technique⁸⁸, *i.e.*, defining an initial curve in the image which evolution is controlled by a cost or energy function⁸⁹. This cost function has to be minimized and should be chosen to reflect the task at hand.

Classical PDEs-based methods for segmentation include level sets, active contours, and morphological geodesic active contours.

Level set approaches represent the studied evolving contour as a signed function, where its zero-level is the actual contour⁹⁰. Level set approaches are particularly useful when studying

shapes with changing topologies. A popular level sets based approach is the Morphological Chan-Vese algorithm⁹¹ which is designed to segment objects with no clearly defined boundaries.

Active contours, also called snakes, define the energy function as a sum of internal and external energies⁹². The external energy function should be minimum at the object borders. This can be achieved by matching its lowest values to the positions with the highest gradient. While the internal energy function is dependent on the shape of the object, *e.g.*, granting high energy to elongated contours. Segmenting an image with active contours relies on finding a trade-off between the two energy functions, *i.e.*, between a shape regularity and its contours.

Another popular PDEs-based method is morphological geodesic active contour, a method where the active contour is built and adjusted at each iteration based on the application of morphological operations (*e.g.*, dilations and erosions) and geodesic distances. This method is particularly useful when the contours of the objects are noisy, cluttered, or unclear.

2.1.5 Watershed-based methods

Watershed-based methods are a subtype of region-based segmentation algorithms, introduced by Digabel *et al.*, and are very often used for cell segmentation^{93–95}. These methods view images as topographic landscapes with ridges and valleys⁹⁶. The 2D images are represented in a 3D space where the third dimension, *i.e.*, elevation, is defined by the gray scale intensity level. This visualization creates catchment basins in the image. For each local minimum, a basin is made from all the points which path of steepest decent terminates at that minimum. The watershed-based methods will then separate the basins in the image, identifying distinct regions.

In biomedical images, where signal-to-noise ratio is low and images are complex, watershed-based methods tend to over-segment the images⁹⁷ creating more region/basins/cells than necessary. Solutions forcing the merging of adjacent basins have been proposed to counter the

over-segmentation problem^{98–100}. Segmentation can also be improved by using seeds at the approximate locations of cell centers to initiate a marker-controlled partitioning process.

2.1.6 Marked point processes

A marked point process (MPP) is a representation of data in a k -dimensional Euclidean space¹⁰¹ that allows the detection of predefined objects by defining their geometrical description and their spatial distribution. MPPs are particularly useful in the case of object detection in biomedical images. Indeed, biomedical images often contain multiple object types, *e.g.*, cells of interest, proteins, vesicles, and noise, and so cannot be modelled simply as foreground and background, but additional information of the specific targeted objects is needed. By modeling only the objects of interest, MPPs bypass this issue¹⁰².

MPPs can be used to detect both simple-shaped objects and ones with complex morphology, and thus are very useful in biomedical images, *e.g.*, detecting pleomorphic nuclei of cancerous cells¹⁰³. MPPs are also particularly useful when dealing with noisy non-uniform images containing irregular unknown structures¹⁰⁴, or when the spatial dependence between structures is of interest^{105,106}.

2.1.7 Clustering-based methods

Image segmentation by clustering is the process of grouping similar pixels together to form a single segment based on their level of similarity. Like region-based segmentation methods, clustering-based methods can be agglomerative or divisive¹⁰⁷. In cell segmentation, agglomerative algorithms, in which cluster sizes increase iteratively to finally form the full cells, are more common.

A popular agglomerative clustering algorithm is K-means^{108–113}, an unsupervised algorithm that partitions the image into K -number of clusters based on K -centroids such that the distance between elements of the same group is minimized. K-means has been used for cell

segmentation, but requires additional steps, *e.g.*, ellipse fitting, hole filling, or separation of merged cells, to ensure that cell shape is respected. It is also limited by varying intensity levels within a same cell leading to over-segmentation.

All of these traditional methods can be combined into pipelines^{108,114,115} for cell segmentation, improving their results step-by-step. They can serve as pre-processing to more advanced techniques to improve image quality before segmentation, *e.g.*, enhancing edges detectability, reducing the noise in the image while preserving small structures, defining the region-of-interest, and reducing the intensity variation within it, or all the above. Or they can be used as post-processing steps to improve segmentation results.

Fig. 6 shows the output of each traditional method discussed above applied to the same image representing coins. Overall, the Laplacian of a Gaussian method and active contours gave the least satisfactory results in identifying the coins. Indeed, Laplacian of Gaussian failed to detect all coin contours, whereas the active contours method was limited by the initial curve chosen to find the object and its result were muddled by neighboring shapes. The other tested methods successfully identified the coins with different levels of precision at the border with the background. However, the best approach here was the watershed-based method, which identified each individual coin on the image as a separate shape.

In the following work, pre- and post- processing will refer to the steps applied to the image before or after the detection of cells.

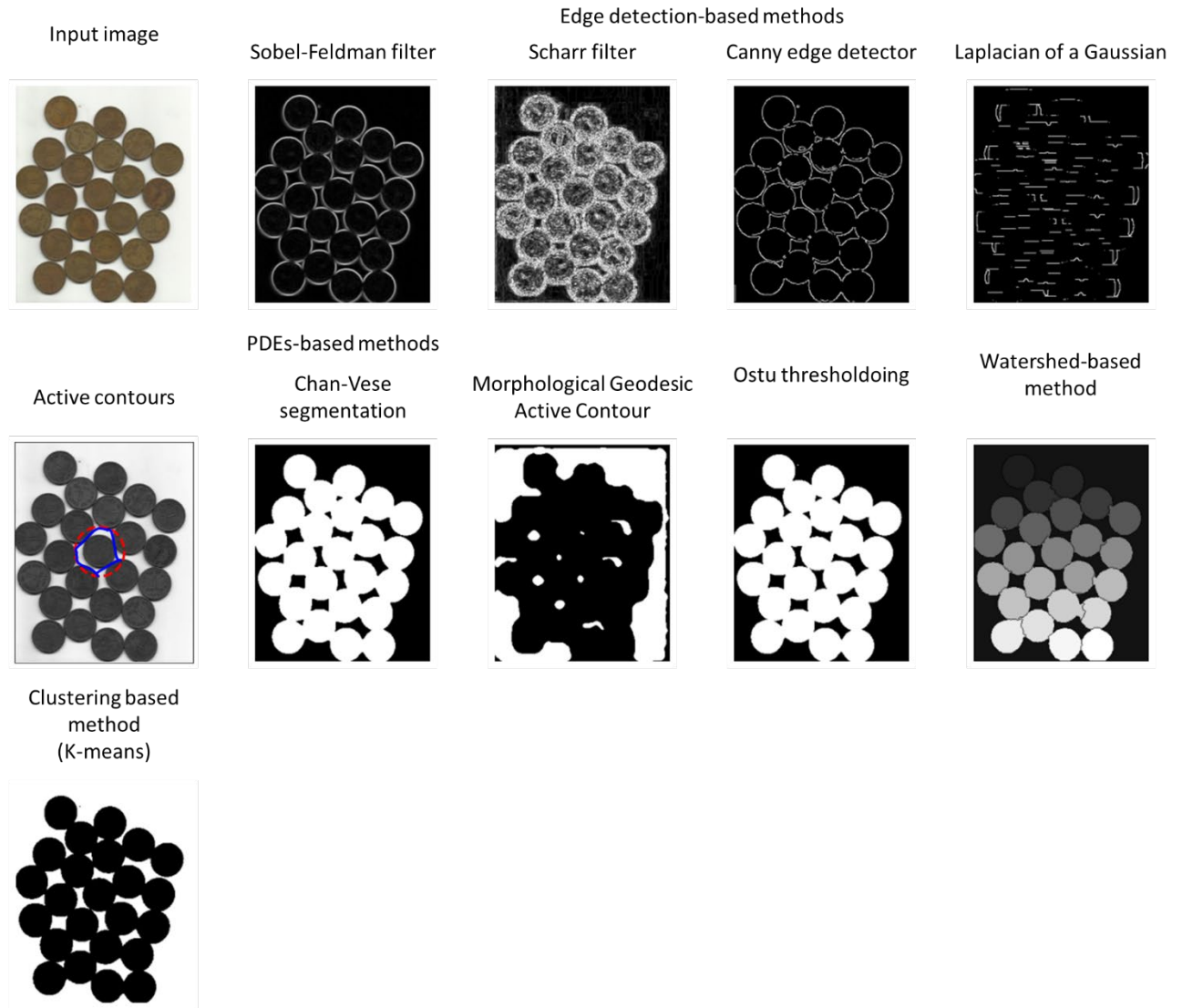


Fig. 6 Results of applying different traditional methods for segmentation on a same image.

2.2 Deep learning-based approaches

Deep learning-based methods have been employed to analyze biomedical images, gaining increasing popularity fueled by their ability to use considerable amounts of data to uncover patterns, undiscernible to the human eye and brain, achieving great accuracy and power while maintaining flexibility and generalizability. Deep learning allows for the extraction of high-level features from the data in an incremental hierarchical fashion.

Most deep-learning based methods are based on CNNs, but architectures based on fully convolutional neural networks, *e.g.*, U-net, have gained in popularity due to their great results in segmenting images. In 2015, GANs were introduced and later described by Yann LeCun as “the most interesting idea in the last 10 years in machine learning”. GANs and cycle-GANs have great potential in biomedical imaging, opening the door to higher quality images by increasing their resolution, and to the generation of better-quality synthetic images for data augmentation. Indeed, common problems encountered in biomedical image acquisition and analysis, are the poor quality and the lack of data and labelled data required for the training of deep learning models.

In the following paragraphs, we will briefly discuss CNNs, U-net, GANs and cycle-GANs. This is nowhere near an exhaustive review of deep learning methods used in the segmentation of biomedical images, nor a comprehensive review of the application of the four aforementioned deep learning architectures in this field but rather an introduction to important methods that will be employed in the following chapters of this thesis.

2.2.1 CNNs

The most used deep learning algorithms are based on CNNs, which were introduced by LeCun *et al.* in 1989¹¹⁶. Contrarily to conventional machine learning methods, they do not require manual parametrization of hand-crafted features, can be parallelized using GPUs/CPUs, can process huge amounts of structured data picking up unknown patterns, outperforming other methods and becoming the state-of-the-art in many computer vision applications including cell segmentation.

The architecture of a typical CNN is a multi-layered-feed-forward neural network composed of a stacking of sequential layers. First, an input layer, followed by hidden layers where knowledge is processed, and finally an output layer. This sequential structure allows the CNN to process

data and learn features. Hidden layers are followed by activation layers that add non-linearity into the CNN, and sometimes pooling (down/up scaling) layers where neighboring pixel values are aggregated together using a permutation invariant function¹¹⁷.

CNNs are resilient and translationally invariant. They are based on a kernel which slides across the image detecting features on the image, independent of their location, as opposed to fully connected networks where all nodes are connected to each other. Thus, CNNs have fewer features and are less computationally voracious.

CNNs have often been used for cell segmentation on biomedical images^{118,119} by sliding through the image using a window and classifying each pixel in the image individually. However, this method leads to overlapping between neighboring patches and therefore repeated computations. To avoid these unnecessary operations, CNNs were modified to return a likelihood map, rather than one output for a single pixel. The resulting architecture is called a fully convolutional neural network¹²⁰ (fCNN). Unfortunately, fCNNs contain pooling layers which cause a loss of resolution. To solve this issue, Ronneberger *et al.* proposed the U-net architecture¹²¹.

2.2.2 U-net

U-net is a fCNN, made of two symmetrical paths forming a U-shape (see Fig. 7). The first path is a contracting one which captures the context information and is an encoder network. It is made of a succession of 3*3 convolutions followed by a rectified linear unit (an activation function defined as the positive part of its argument) and 2*2 max pooling for down sampling. Each down-sampling operation doubles the number of feature channels. The contraction reduces the spatial information while augmenting the feature information. The second path is an expanding one, *i.e.*, a decoder network, and captures localization information. It consists of a series of up-sampling followed by a 2*2 up-convolution, which halves the number of feature

channels, concatenation with the cropped feature map from the symmetrical contracting path and two 3*3 convolutions, each followed by a rectified linear unit. The large number of features in the expanding path allows the network to propagate context information through the network to higher resolution layers.

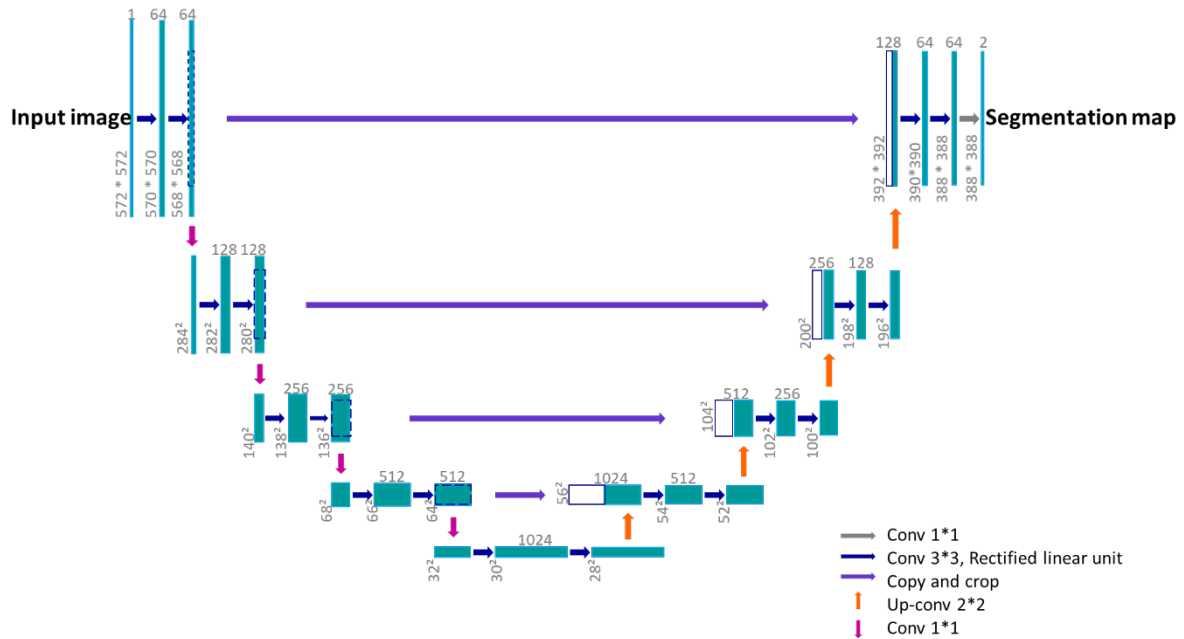


Fig. 7 U-net architecture. A teal box represents a feature map. The number of channels is indicated on top of the feature map box. The image size in pixels is indicated on the lower left side of the box. Boxes with dark blue contours represent copied feature maps. The arrows indicate the different operations.

U-net is very useful in image and cell segmentation, as its use of skip connections allows for the incorporation of both high- and low- level image features in the segmentation result, therefore requiring less training data, often an issue in biomedical imaging. This has led to the widespread use of U-net with all types of imaging modalities¹²² often outperforming other methods.

Other versions of U-net have been developed for different applications. Attention U-nets to provide the network with the ability to focus on specific objects¹²³, Inception U-net to capture objects of different sizes and shapes through the use of filters of multiple sizes in the same layer

in the network¹²⁴, Residual U-net which uses ResNet¹²⁵ blocks in order to solve the vanishing gradient problem by adding skip connections between successive layers¹²⁶, and Adversarial U-nets a type of conditional GANs¹²⁷ where each network is a U-net.

2.2.3 GANs and cycle-GANs

GANs and cycle-GANs are two generative deep learning algorithms used respectively for paired and unpaired image translation.

Generative models are a subtype of deep learning methods used to uncover patterns in a given dataset A and learn its architecture, and then create similar but novel and original content similar to A using dataset B .

The key to these two networks is their indirect implicit¹²⁸ adversarial approach to generative modeling. Indeed, one part of the models is generative, while the second part is discriminative. The generative part aims to create realistic target candidates without explicitly learning the data likelihood function, while the discriminator evaluates them. The contest in place between the two parts of these models allows for the improvement of both parts.

Cycle-GANs have the specificity of training two pairs of generator/discriminator models for unpaired image-to-image translation, introducing a cycle-consistency loss to train both pairs simultaneously. This significantly reduces the data requirement, *i.e.*, no requirement for dual image acquisition of a single sample which often is not possible due to the destructive nature of certain methods.

Cycle-GANs learn the special characteristics of one image collection A and determine how these characteristics could be translated into a second image collection B , all in the absence of any paired training examples.

A cycle-GAN learning the translation between these two image domains A and B, consists of two generator networks G_A and G_B . The first is responsible for transforming images from domain A to domain B, while the second transforms images from domain B to domain A.

Each generator is made of three sections: an encoder, a transformer, and a decoder¹²⁹. The encoder is made of three convolution layers. The resulting activation is then passed to the transformer, a series of nine residual blocks. It is then expanded again by the decoder, which uses two convolutions to enlarge the representation size, and one output layer to produce the final image, as seen in Fig. S3.

The generators are associated with discriminator networks, D_A and D_B . D_A tries to distinguish between real images from domain A and fake images generated by G_A , and vice versa for D_B .

The discriminators are PatchGAN networks, a type of fully CNNs that look at a patch of the input image and output the probability of said patch being *real*. By looking at patches and not the full image, the discriminator is more computationally efficient and more effective, allowing the discriminator to focus on more surface-level features of the image.

Cycle-GANs are constrained by an adversarial loss, which encourages the generators to produce images that their matching discriminators cannot distinguish from real images. This loss function is often a least squares loss¹³⁰, defined as $Adversarial\ loss(G_A, D_B, I_A) = \frac{1}{m} \sum_{i=1}^m (1 - D_B(G_A(I_{A_i})))^2$, where m is the number of pixels in image I_A , an image from domain A.

Generators are constrained by a cycle-consistency loss ensuring that $G_B(G_A(I_A)) \approx I_A$. This loss helps the model learn meaningful translations and prevents it from producing arbitrary results, it is often an L1-norm or a summed absolute difference in pixel value between $G_B(G_A(I_A))$ and I_A , and $G_A(G_B(I_B))$ and I_B , where I_A and I_B are images from domains A and B respectively.

To ensure that the generator does not change an image unnecessarily, an identity loss is also introduced. It enforces that $G_A(I_B) \approx I_B$.

GANs and cycle-GANs are gaining traction in cell and biomedical images segmentation, offering a solution to the limited and sometimes biased data sets and opening the door to the possibility of domain-translation.^{131–134}

Overall deep learning methods increase segmentation accuracy, leveraging huge amounts of data and heavy computational power, but with a decrease in interpretability. Unfortunately, in the case of biomedical imaging, databases sizes are limited, particularly ones with annotations, hindering the performance of deep learning-based methods, and data augmentation with generative models may not be enough.

2.3 Multi-task learning

A promising more recent development in deep learning is multi-task learning. Multi-task learning performs multiple related but not identical tasks in parallel and leverages information from all of them to improve the overall performance. This idea is based on the hypothesis that related tasks jointly learn better than performing them individually¹³⁵, especially when training data (labelled or unlabeled) availability is limited, which is often the case in biomedical imaging¹³⁶. Multi-task learning is therefore a solution to alleviate the data sparsity issue by *reusing* available knowledge from one task to another and reducing the cost of manual labeling¹³⁷. It is also a way of reducing the overall memory required to perform the tasks, as it avoids the repeated learning of shared features between tasks¹³⁸. Additionally, it can hypothetically learn more generalized features by averaging the noise patterns in the performed tasks and prioritize the more important features which can be harder to distinguish when performing one task alone.

Multi-task learning can be performed on supervised or unsupervised tasks, regression or classification, requiring labelled or unlabeled data. It has been used in the case of biomedical images segmentation with great success^{139–145}, often outperforming other state-of-the-art methods.

Multi-task learning approaches can be split into two architectures: hard¹⁴⁶ and soft¹⁴⁷ parameter sharing. The first describes multi-task learning architectures where tasks, often neural networks, share hidden layers. This architecture works best when tasks are closely related to each other. The second architecture describes multi-task learning approaches where tasks have their own hidden layers, and share parameters or information through other ways, such as their loss functions or their constrained layers.

Multi-task-based models tend to perform well on domain adaptation and generalization and are therefore less data *dependent*. On the other hand, such adaptability may cause lower pixel-level segmentation precision.

All of the deep-learning methods presented here are task-specific, *i.e.*, the data used for training and testing match the intended goal of the model. In recent years, we have seen a paradigm shift in deep learning-based approaches with the rise of foundation or generalization models¹⁴⁸. These algorithms have been trained on millions of data points and can be adapted to a wide range of tasks with no or minimal retraining. This is often made possible with: (1) support self-supervising tasks generating the labels required for training through transfer learning, *i.e.*, taking the knowledge from one surrogate task and fine-tuning it to downstream another. (2) Scale, *i.e.*, a combination of improvements in computer hardware and the Transformer model¹⁴⁹ architecture, which uses self-attention to determine which features are important in the input. And (3) the availability of larger amounts of data and possibility to store it¹⁵⁰. In April 2023, Meta launched the Segment Anything Model trained on 11 million images and their matching 1.1 billion masks, mostly generated through self-supervision. Though this incredible amount of

images, this foundation model has learned a *notion* of what objects are and is therefore capable of segmenting previously unknown and unseen objects. This is called zero-shot generalization. The methods proposed in this thesis will be task-specific and follow the rough division of cell segmentation methods. We will compare traditional methods, to deep learning-based approaches, and to a multi-task learning architecture for the automatic segmentation of keratinocytes on RCM images.

Chapter 3 A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

RCM allows real-time *in vivo* visualization of the skin at cellular level. The study of RCM images can provide information on the structural properties of the epidermis and upper layers of the dermis. The first method we propose to streamline the extraction of these properties is based on the enhancement of the structural features (membranes) visible in RCM images of the SG and SS, where individual keratinocytes are clearly observable, using a filter adapted to fine and elongated structures. Due to the fact that SG cells are bigger than SS cells and that this approach is based on the detection of their morphological features, a different set of parameters will be used on RCM images of each layer.

This method will be referred to as the Full Image Analysis Pipeline (FIAP). Both the method itself^{151,152} and the biological insights¹⁵³ derived from its application to over 5000 clinical RCM images have been published.

The obtained results were compared to a manually obtained ground truth of cell positions and achieved an accuracy (precision and recall) on par with expert graders.

The proposed FIAP will be compared in this chapter to various machine learning-based approaches: the Cell Cutter¹⁵⁴ algorithm and different configurations of the U-net¹²¹

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

architecture for image segmentation. In a later chapter, it will be compared to a novel multi-task learning architecture.

3.1 Full Image Analysis Pipeline (FIAP)

The proposed method is a 3-step approach to obtain cell positions from RCM images of the SG and SS: (a) We first identified the region of interest containing the epidermal cells (in blue in Fig. 8), (b) we then segmented the individual cells (in red in Fig. 8) in the identified tissue area using tubeness filters to highlight membranes, (c) and we finally used prior biological knowledge on cell size to process the resulting detected cells, removing cells that are too small and reapplying the used tubeness filters locally on detected regions that are too big to be considered as a single cell.

All steps were applied to a framed image, where an 11.6% frame was synthetically removed to counter for the vignetting problem in microscopy, *i.e.*, uneven, and poorer illumination at the image periphery due the microscope light path.

3.1.1 Identification of the region of interest

The FIAP starts with the detection of the region of interest (ROI). This allowed us to focus on only the critical portion of the image which contains the objects of interest (keratinocytes), thus preventing false alarms and saving computational time and power by not processing non-critical areas (containing objects that are either not of interest or of low quality due to noise or heterogeneity) and background (void of cells).

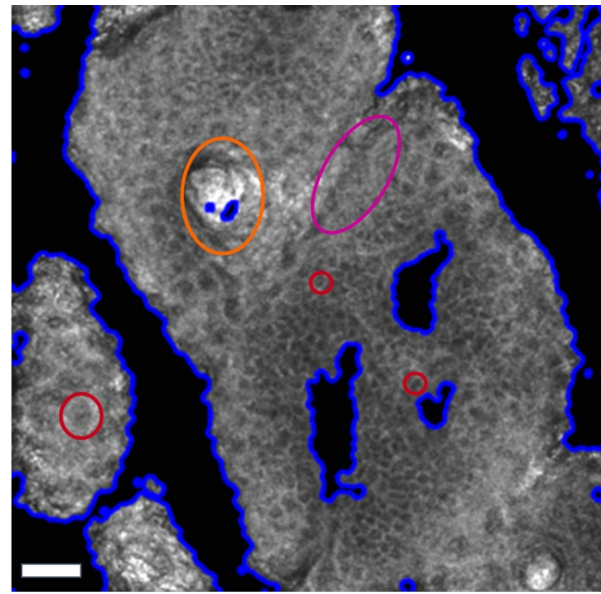
The dark background is distinguished from the tissue area (see Fig. 9) by first identifying the tissue islands borders. Indeed, skin micro-relief lines surround keratinocytes⁶ and appear in RCM images as dark regions enclosing islands of cells. In order to identify the area representing these furrows, we used a morphological geodesic active contour algorithm¹⁵⁵, also known as

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

the Snake algorithm, which applies morphological operations (dilations and/or erosions) to the image in order to detect visible contours based on their intrinsic geometric measures, even if said contours are noisy, fuzzy, cluttered or partially hazy, by minimizing the energy function assigned to the surface S^{156} (Eq. 1),

$$E(S) = \iint g(I)(S(a))da \quad (1)$$

where da is the Euclidean element of area, $S(a)$ is the surface area, and where the region of interest on the image I is defined by $g(I) : \mathbb{R}^d \rightarrow \mathbb{R}^+, x \rightarrow g(I)(x)$ and typically is $g(I) = \frac{1}{\sqrt{1+\alpha|\nabla G_\sigma * I|}}$ with G_σ * a gaussian filter with standard deviation σ and α a user defined weight.



○ Epidermis cell ○ Non-informative areas
○ Epidermis tissue ○ Bright areas

Fig. 8 RCM image of the *stratum spinosum* of minimally pigmented skin, Fitzpatrick phototype II. In blue, the border between tissue and background formed by micro-relief lines. In pink, non-informative areas. In orange, bright spots. In red, epidermal cells. RCM, Reflectance confocal microscopy. Image contrast was adjusted for better visualization. Scale bar = 50 μm .

The morphological snake algorithms are faster and more stable than their standard (geodesic) active contours counterparts. The latter use partial differential equations and are not easily

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

trapped by local minima but require pre-processing of the image to highlight the contour. To initiate the snake algorithm, we used the binarized distance transform of the opening of the Otsu binarized vignettted image, *i.e.*, a derived representation of the binarized image, where the value of each pixel is the nearest distance to an approximate background (nearest zero pixel).

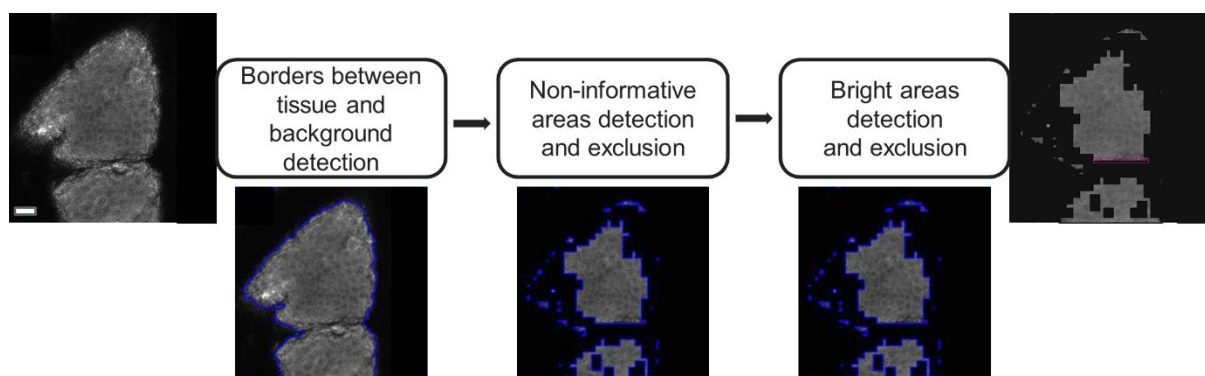


Fig. 9 Steps of the identification of the ROI. The borders between the tissue and the background are identified using a Snake algorithm. The ROI was then refined using a Support Vector Machine algorithm trained to detect the non-informative areas, and a succession of morphological operations to remove bright spots in the RCM image. In blue is the border of the region of interest. In pink is an area that should have been removed through the different ROI identification steps but was not. Image contrast was adjusted for easier visualization. Scale bar = 50 μm . ROI, region of interest; RCM, Reflectance confocal microscopy.

After identifying the background formed by micro-relief lines on RCM images, the next step was to identify non-informative areas. These areas are part of the tissue but are of low quality (see Fig. 8) and individual cells are not clearly and easily distinguishable, by neither the expert manual grader nor any algorithm. To discriminate these non-informative areas from critical regions clearly containing cells, a support vector machine (SVM) was trained to perform a texture classification using four features of the grey level co-occurrence matrix¹⁵⁷ (GLCM). SVMs are supervised machine learning models for classification and regression which aim to find a hyperplane that distinctly classifies the data points such that the distance between data points of each class is maximized. The GLCM is a measure of the frequency of occurrence of two neighboring pixels combination in an image, this supposes that the texture information is

comprised in the spatial relationship between neighboring grey scale pixels. Since GLCMs are often sparse, they are summarized in 22 Haralick features (named after their inventor). We only focused on 4 of these features that successfully discriminate between informative and non-informative areas (see Fig.). The remaining 18 features did not distinguish between informative and non-informative areas. In the formulas below, P is the GLCM histogram for which to compute each feature, for a gray level j distant from a gray level i . These features are:

(a) Homogeneity¹⁵⁸ (Eq. 2);

$$homogeneity = \sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (2)$$

which measures the closeness of the GLCM distribution to its diagonal, reflecting correlation by measuring the uniformity of the non-zero entries in the GLCM¹⁵⁹. The more pixels with similar grey level values, the closer the GLCM to its diagonal, reflecting a higher homogeneity value.

(b) Contrast¹⁵⁸ (Eq. 3);

$$contrast = \sum_{i,j=0}^{levels-1} P_{i,j} (i - j)^2 \quad (3)$$

which measures the local variations in the GLCM.

(c) Dissimilarity¹⁵⁸ (Eq. 4);

$$dissimilarity = \sum_{i,j=0}^{levels-1} P_{i,j} |i - j| \quad (4)$$

which measures the similarity between pixels by measuring the variation of grey level pairs in the image. It ranges from 0 to 1 and is at its highest when the differences between the pair are at their maximum. Contrast and dissimilarity are close in meaning, with only a difference in

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

weight. Contrast grows quadratically while dissimilarity does not (see Eq. 3 and Eq. 4) (see Fig. 10).

And (d) energy¹⁵⁸ of the GLCM(Eq. 5),

$$energy = \sqrt{\sum_{i,j=0}^{levels-1} P_{i,j}^2} \quad (5)$$

Also known as angular second moment or uniformity, which measures the signal (texture) uniformity within the area, *i.e.*, local homogeneity. The higher the energy ([0,1]), the more homogeneous the image. Indeed, as the GLCM is a measure of the frequency of occurrence of two neighboring pixels combination in an image, if $energy = 1$, then the image is constant.

The GLCM was calculated on 50x50 patches, and prediction of informative vs. non-informative area was computed per patch. A sliding window with a 32 step was used to parse through each image.

The SVM was trained on 6 images to distinguish between informative and non-informative areas as they are able to separate the two (see Fig. 10). A total of 50 informative patches and 18 non-informative patches were selected for the training of the model, and all 4 Haralick features were computed for each patch. The SVM achieved a 66.7% precision (the fraction of correctly detected regions among all the detected regions) and a 100% recall (the fraction of accurately detected regions among all regions defined in the ground truth).

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

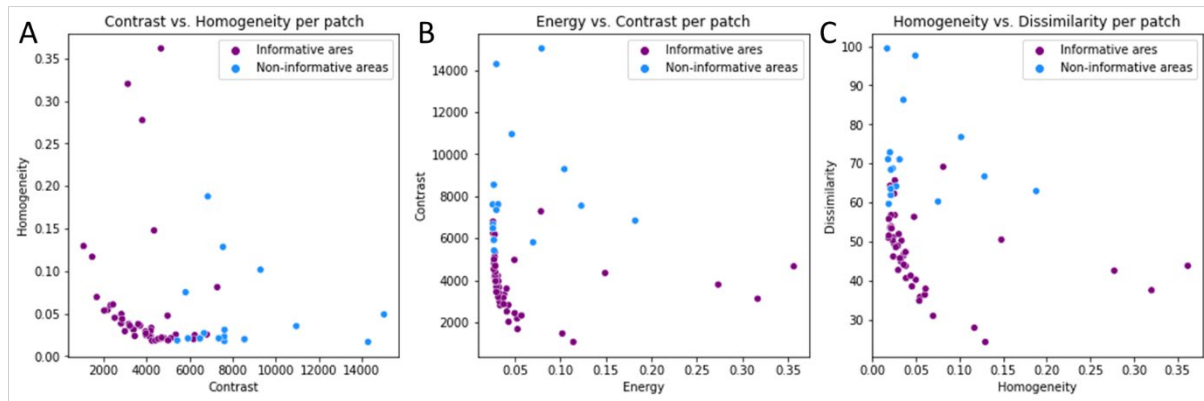


Fig. 10 Four Haralick features were used to train an SVM to distinguish between non-informative and informative areas in RCM images. **(A)** Contrast vs. Homogeneity per patch. **(B)** Energy vs. Contrast per patch. **(C)** Homogeneity vs. Dissimilarity per patch. SVM, Support Vector Machine; RCM, reflectance confocal microscopy.

The third step in ROI identification was to remove any bright spots in RCM images (see Fig. 8). These may be caused by the keratin in hair shafts, the presence of corneocytes in images of the SG and SS, or from mottled pigmentation. To do so, we applied a succession of dilations and erosions on the image where the background and non-informative areas had already been removed, and which had been convolved by a Gaussian filter and binarized with a manually determined binary threshold.

Hyper-parameters of each step are detailed in Table 6.

3.1.2 Identification of individual cells

The second step of the FIAP is to identify the cells within the ROI. We started by applying a median filter to the masked image to remove the noise, followed by a local normalization which uniformizes the mean and variance of the image locally to correct uneven illumination and shading artifacts (see Fig. 11). We then applied a Sato vesselness enhancement filter¹⁶⁰ to this intermediate output to highlight white continuous ridges in the image, *i.e.*, bright cell membranes. This filter performs an analysis of the second derivatives based on a tube model¹⁶¹. It is defined as:

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

$$F = \begin{cases} \lambda_c e^{\frac{-\lambda_1^2}{2(\alpha_1 \lambda_c)^2}} & \text{if } \lambda_1 \leq 0, \lambda_c \neq 0 \\ \lambda_c e^{\frac{-\lambda_1^2}{2(\alpha_2 \lambda_c)^2}} & \text{if } \lambda_1 > 0, \lambda_c \neq 0 \\ 0 & \text{if } \lambda_c = 0 \end{cases}; \quad (6)$$

Where λ_i are the Hessian matrix eigen values such that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. To obtain a high response in tubular structures, the Sato filter introduces an asymmetric formulation based on the sign of λ_1 to reduce noise. The α_i parameters control this asymmetrical strength with $\alpha_1 < \alpha_2$, and $\lambda_c = \min \{-\lambda_2, -\lambda_3\}$ ¹⁶². In this framework, when $\lambda_2, \lambda_3 < 0$, the eigen vector associated with λ_1 corresponds to the direction of the putative vessel, and the eigen vectors associated with the other two eigen values form the basis for the cross-section of the vessel.

The filter parameters were empirically chosen to approximate the width and length of a keratinocyte membrane in the SG and SS.

To the output of Sato's vesselness filter, we applied a median filter and local normalization with respect to the ROI mask (see Fig. 11). We then further enhanced the vessel-like structures by applying a Gabor filter which convolves the previously obtained image by a windowed sinusoidal signal of varying frequencies and orientations (see Fig. S1), modulated by a Gaussian envelope¹⁶³, resulting in an image where the highest response is located at edges and points where a change in texture is observed. The sinusoidal signal used in the Gabor filter has both a real and an imaginary component forming the complex equation:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{x'}{\lambda} + \psi\right)\right) \quad (7)$$

With $x' = x \cos\theta + y \sin\theta$ and $y' = -x \sin\theta + y \cos\theta$.

Here λ is the wavelength of the sinusoidal component and controls the width of the Gabor function strips, θ is the orientation of the normal to the parallel strips of the Gabor function, ψ is the phase offset of the sinusoidal function, σ is the standard deviation of the Gaussian

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

envelope and governs its size, and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function which controls its height.

The image obtained after application of the Gabor filter was equalized with a histogram equalization to adjust its contrast by spreading out the intensity range of the image, followed by a Gaussian adaptive thresholding which locally changes the binarization threshold over the entire image to account for local changes in contrast and brightness, this assumes that local areas in the image are more likely to be similar with respect to contrast and illumination. The local binarization threshold is defined as the gaussian-weighted sum of neighboring values. A connected-components analysis/labelling was then applied to the binarized image and its inverse, to detect and remove small blobs due to noise, and to close small holes in the membranes due to image graininess. This is done by building a graph from the image which is used to identify the positions belonging to the same component/label and their respective size. Finally, we skeletonized this last result into 1-pixel wide centerlines while keeping the essential structure, topology and Euler characteristics of the detected regions¹⁶⁴, and then pruned unwanted parasitic spurious branches from the skeleton, following a method adapted from the MATLAB routine `bwmorph`.

Hyper-parameters of each step are detailed in Table 6.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

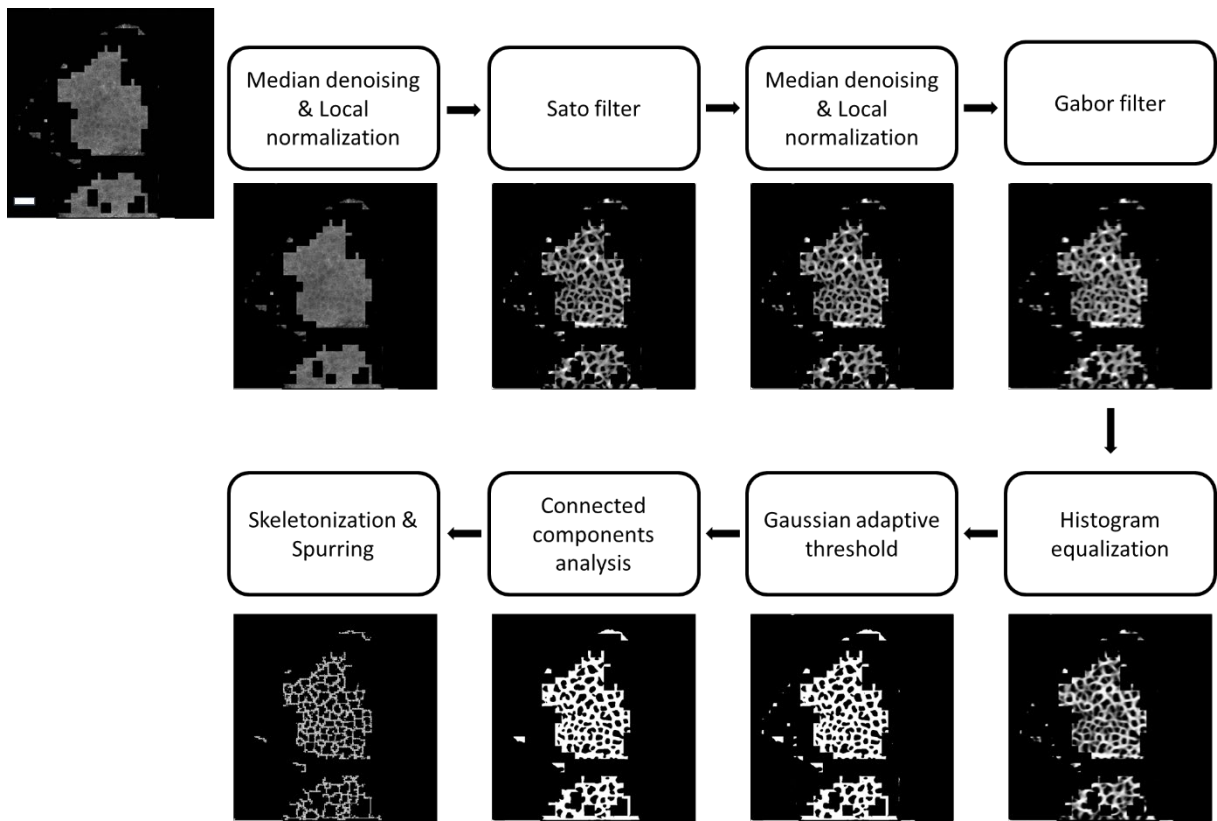


Fig. 11 Image processing steps for the identification of individual cells. A median filter and a local normalization were applied to the image with the ROI mask, followed by a Sato filter. Its output was filtered with a median filter and locally normalized, and a Gabor filter was applied to it. A threshold was applied on the output after histogram equalization and small blobs were removed with a connected components analysis. The result was then skeletonized, and spurious branches were removed. Image contrast was adjusted for easier visualization. Scale bar = 50 μm . ROI, region of interest.

3.1.3 Post-processing

The pruned skeleton was further cleaned with a morphological closing to fill small gaps. Individual contours, *i.e.*, cells, were detected on the clean skeleton (see Fig. 12). We then set an empirical threshold, based on the knowledge of cells size in the different epidermal layers, on the minimum cell area and removed contours that were smaller than said threshold ($area < 100 \text{ pixels}$ for contours detected on RCM images of the SG and $area < 50 \text{ pixels}$ for contours detected on RCM images of the SS). Long contours, often detected at the frontier between background and ROI, were also removed based on an empirical threshold on contour

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

eccentricity *i.e.*, $eccentricity > 0.85$ (see Fig.12-A). The remaining detected contours were then separated into 2 groups: (1) large contours with an $area > 1000 \text{ pixels}^2$ for SG and $area > 120 \text{ pixels}^2$ for SS, and (2) small acceptable contours with an $area \leq 1000 \text{ pixels}^2$ for SG and $area \leq 120 \text{ pixels}^2$ for SS. On each area of the original vignettted masked image delimited by what has been defined as a large contour, we reapplied a Sato filter locally with different, more refined, parameters for more precise membrane enhancement and detection. This local output is then binarized, with an Otsu thresholding for RCM images of the granular layer, and with a Gaussian adaptive thresholding for RCM images with a majority of SS cells. Double connected-components labelling was again performed to get rid of contours and holes formed by noise. The resulting image was skeletonized, and the contours were detected on the obtained skeleton. On images of the SG, obtained contours with an $area \leq 110 \text{ pixels}^2$, were merged back with their neighbors. We considered that these new contours were too small to be actual granular keratinocytes, and therefore were an artifact of the second iteration of Sato filtering (see Fig. 12-B). On the contrary, the second round of Sato filter with different, more refined, parameters may still have failed to detect cells on RCM images of the SS, as their signal may be too small to be picked up. To counter this issue, we chose to fit ellipses within the large SS contour if the second iteration still failed to divide the contours into smaller ones. The size of the ellipse was defined as the local median size, which was defined by the size of the minor and major axes of the ellipse, of the surrounding *correct* contours.

All of the newly obtained cells were then combined with the previously found small acceptable contours and their centers were detected.

Hyper-parameters of each step are detailed in Table 6.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

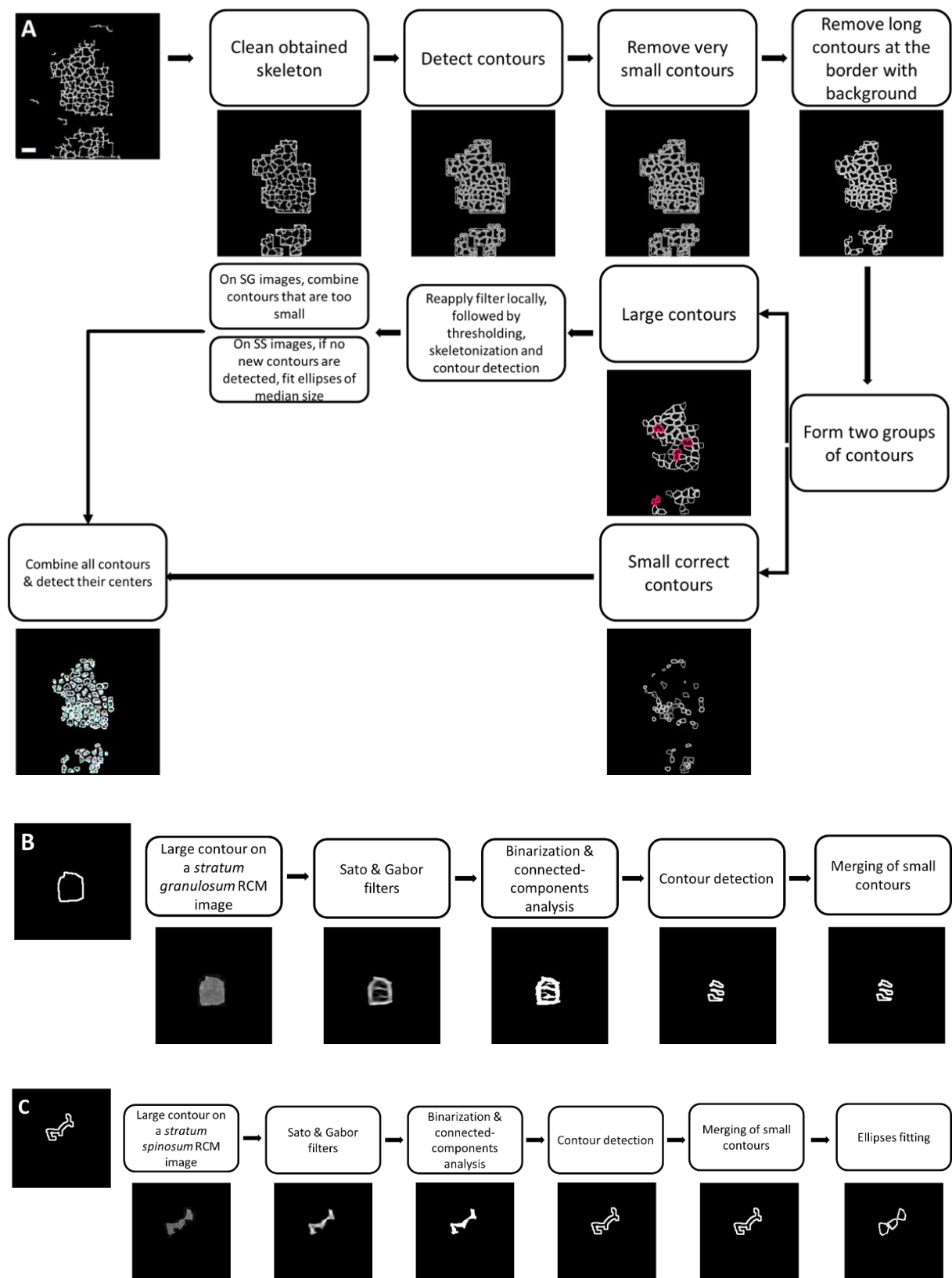


Fig. 12 Post-processing steps. (A) The skeleton obtained after the previous step was cleaned, and contours were detected. Small contours were removed, as well as long contours close to the border with the background. The remaining contours were divided into two groups: small

and big contours. Big contours were filtered again to improve the detection locally. The new resulting contours were then combined to the small contours and their centers were detected. In pink, some contours where two cells were merged are highlighted. **(B)** Example of large contours improvement for a *stratum granulosum* image. **(C)** Example of large contours improvement for a *stratum spinosum* image. Image contrast was adjusted for easier visualization. Scale bar = 50 μm .

3.2 Data

In two separate clinical studies, RCM images were collected: (a) on the volar forearm of 80 participants: 60 children, ages 3 months to 10 years, and 20 adults, ages 25 to 40 years, and (b) on the volar forearm and cheek for 80 women, ages 40 to 80 years. The studies were initiated following approval from an independent institutional review board (studies 19.0198 and 20.0022) and in accordance with the Declaration of Helsinki. Study participants or their legal guardian (in the case of children) gave written informed consent prior to the study. Participants were divided into 11 age groups: 0-1, 1-2, 2-4, 4-6, 6-8, 8-10, 20-40, 40-50, 50-60, 60-70, and 70-80 years of age (see Table 3). Age group 1-2 years old was removed from the subsequent analyses due to insufficient data (only 4 participants with usable data).

All participants have minimally pigmented skin, with Fitzpatrick phototypes between I and III (see Table 3). To be eligible for the study, participants had to be in good health, with no history of skin disease, and refrain from applying any products on the observed area on the day of the study.

RCM images were captured using a VivaScope 1500 reflectance confocal microscope. It has a z-resolution of 5 μm and an xy-resolution of 1 μm . The imaging process began at the skin surface and progressed towards the first layers of the dermis. Images were captured in stacks of 82 images, each one of size 1000x1000 pixels corresponding to a 500 μm^2 area.

The proposed method to automatically detect keratinocyte positions were only applied to RCM images of the SG and SS. Therefore, we needed a computerized method for layer classification.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

To do so, we used the hybrid deep-learning method proposed by Kaur *et al.*¹⁶⁵ and previously described in I.3.1.

The development and validation of the algorithm relied on a subset of only 9 images of both the volar forearm and cheek, involving 7 participants aged 5 months to 35 years, and were acquired during 3 different clinical studies. Six (6) of these images have been graded by two experts for comparison purposes. The ground truth used for validation relied on cell centers which were pointed out manually by skin research experts.

When comparing the FIAP to deep learning approaches later in this chapter, the ground truth used for training the models was generated by Voronoi tessellation around the cell centers determined by experts. Voronoi tessellation is a space partitioning method into polygons/cells/regions closest to a predetermined point, called seed. Each point of the 2D Euclidean space is assigned to a cell, such that the distance between the point and the cell seed is less or equal to that of any other seed.

Table 3 Study participants repartition per age group, gender, and Fitzpatrick phototypes.

Age group (yrs.)	0-1	2-4	4-6	6-8	8-10	20-40	40-50	50-60	60-70	70-80
Number of participants	8	10	9	10	10	20	15	15	15	15
Female/Male ratio	5/5	4/6	7/3	6/4	7/3	20/0	15/0	15/0	15/0	15/0
Fitzpatrick phototype I (%)	0	0	0	0	0	5	13.3	6.7	0	6.7

Age group (yrs.)	0-1	2-4	4-6	6-8	8-10	20-40	40-50	50-60	60-70	70-80
Fitzpatrick phototype II (%)	100	90	100	100	80	75	46.7	53.3	53.3	46.7
Fitzpatrick phototype III (%)	0	10	0	0	20	20	40	40	46.7	46.7

3.3 Synthetic images

The automation of cell identification in RCM images presents challenges due to poor image quality resulting from high noise levels and low contrast. Furthermore, assessing the accuracy of any method requires manual labeling to establish a ground truth, which is laborious, susceptible to human error, and subject to variability among different experts. To overcome these challenges and facilitate the parameterization of our automated pipeline, we devised a process to generate completely user controlled synthetic RCM images (see Fig. 13). By creating these synthetic RCM images, we were able to circumvent the limitations posed by poor image quality and the need for manual labeling and the uncertainty in manually set ground truth. These synthetic images provide precise ground truth annotations, in the form of known cell centers, and offer flexibility in terms of the number of images available for analysis.

The process of generating the synthetic images starts by creating a random tissue mask within the intersection of Bezier curves. These are mathematical representations of continuous smooth curves. They are defined by their endpoints, also known as anchor points, and additional points to determine their curvature and direction. Within the created tissue mask, we initiated a

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Voronoi tessellation from random seeds. These seeds, representing cell centers, were first generated on a grid at a minimum set distance, which value depended on the mimicked epidermal layer. The seeds are obtained by simulating a Poisson process including a minimum distance between them (“hard-core process”). Then, only the seeds within the tissue mask were kept and some were randomly shuffled within the mask to imitate the non-uniformity of the epidermal structure. To simulate RCM images, we also added salt and pepper noise and Gaussian noise to the binary skeleton. The resulting image was then convolved with a Gaussian filter and multiplied by a heterogeneous intensity mask to obtain synthetic images that resemble real RCM images.

Voronoi tessellation has been previously used to simulate the geometry of skin cells⁹ and other types of cells^{166–168}, but while these methods create synthetic cells matching the real cell geometry they do not replicate the epidermal tissue topology.

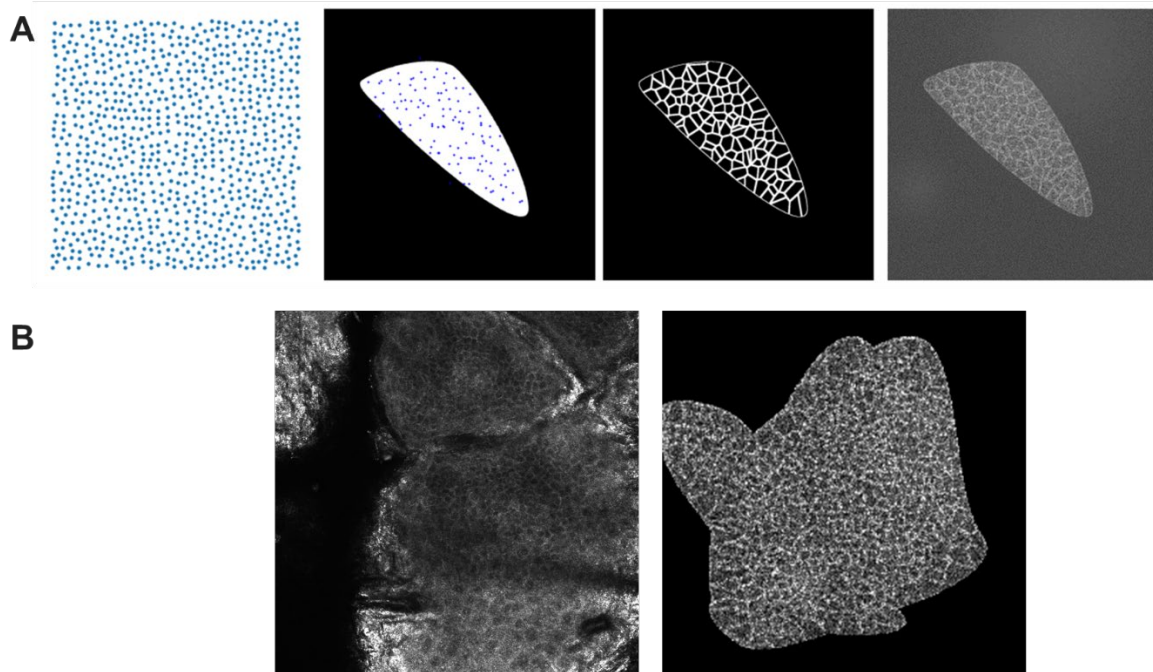


Fig. 13 (A) Process of creation of synthetic RCM image of the *stratum spinosum* created using a hard-core process. First, points set apart with a minimum set distance from each other are created. Second, a tissue mask is created using Bezier curves and only the points within the mask are kept. Third, these points are used as seeds to initiate a Voronoi tessellation. Lastly, different noise levels are added to the image.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

(B) Side-by-side comparison of a real RCM image and a synthetic RCM image. RCM, Reflectance confocal microscopy.

3.4 Accuracy evaluation

A marker-controlled watershed was initiated on the vignettted masked RCM image using the previously determined cell centers (see Fig. 14). The resulting regions were then compared to the ground truth for the 9 RCM images where it had previously been manually determined by expert graders, using the package d-accuracy¹⁶⁹ which evaluates several detection accuracy metrics (see Fig. 14). We chose to focus on two of the most well spread accuracy metrics in biomedical image analysis and in computer vision in general: (a) the precision which corresponds to the fraction of correctly detected cells among all the detected cells,

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

and (b) the recall defined as the fraction of accurately detected keratinocytes among all cells defined in the ground truth,

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Where TP, stands for true positives, FP, stands for false positives, and FN, stands for false negatives.

Since the task here is to compare datapoints to regions, *i.e.*, evaluate if a point in the ground truth (cell center) is within a detected region, the accuracy metrics are not at pixel-level but at object level. This is particularly of interest in the case of RCM images, which are noisy and grainy, and where using pixel-level metrics may lead to low accuracy and therefore to the false belief that the segmentation results are not satisfactory.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

In this case, a TP corresponds to a region obtained from the marker-controlled watershed containing a point from the manually obtained ground truth in the form of cell centers. A FP corresponds to a detected keratinocyte with no matching cell center from the ground truth. A FN is when no keratinocyte was detected where a cell center from the ground truth is.

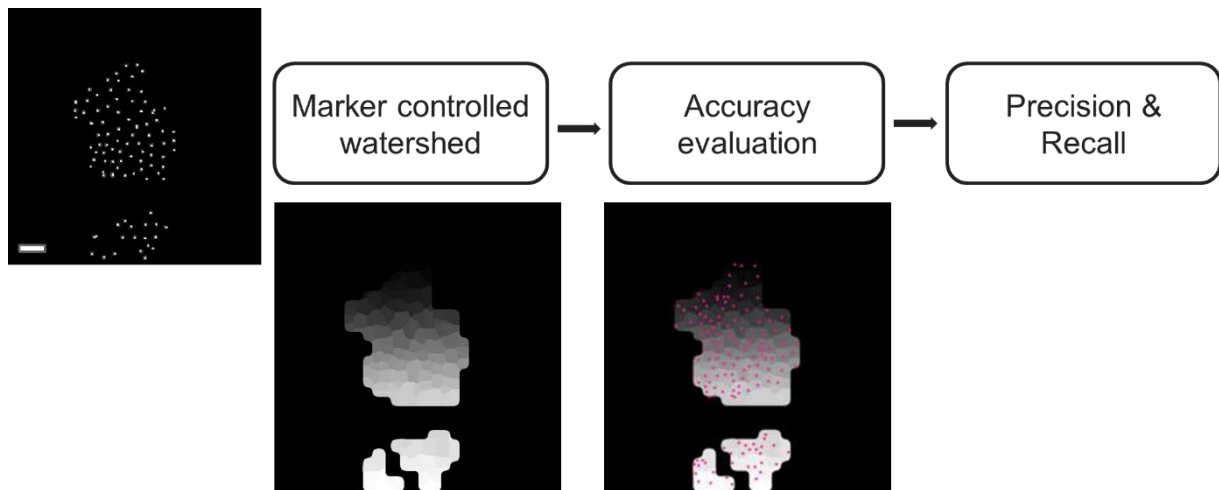


Fig. 14 Detection accuracy evaluation steps. A marker-controlled watershed was applied to the detected cell centers and the resulting labels were compared to the manually detected ground truth, in pink. The returned metrics were precision and accuracy. Image contrast was adjusted for easier visualization. Scale bar = 50 μ m.

The entire FIAP is shown on Fig. 15.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

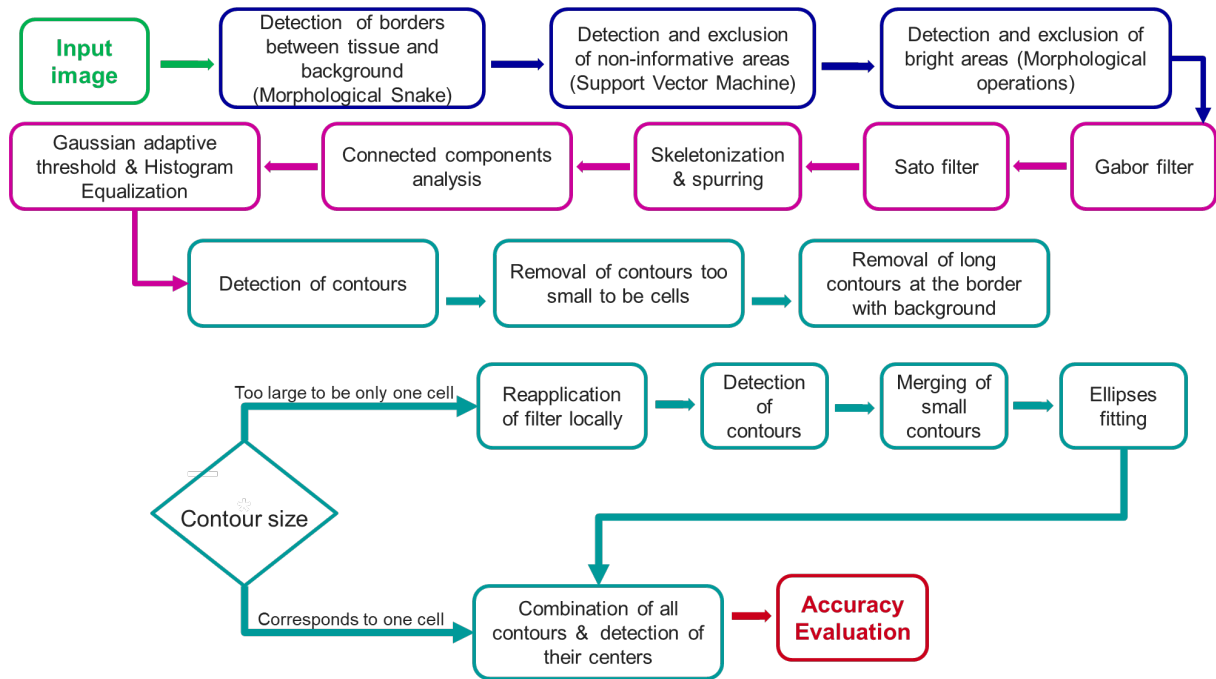


Fig. 35 Diagram of the FIAP used for keratinocytes detection. The sections are color coded as follows: blue, ROI identification; pink, individual cells identification within the ROI; teal, post-processing steps. ROI: region of interest.

3.4.1 Accuracy on synthetic RCM images

To determine a reference to interpret the obtained accuracy, we employed the previously mentioned "hard-core" process to generate multiple synthetic images where randomly generated points were placed within the same correct ROI mask. The number of points chosen matched the number of cell centers in the ground truth, and the distance separating neighboring points was determined as to replicate the epidermal layer represented in the image. We then compared the accuracy of these random detections against the absolute ground truth and obtained a precision and recall of $60 \pm 2\%$, which we established as the lowest accuracy threshold. This threshold serves as a reference to better assess the performance of the algorithm. Application of the FIAP on 4 synthetic images resulted in a median precision of 83.5% with a standard deviation of 6.74%, and a median recall of 92.5% with a standard deviation of 1.22%. These results demonstrate the high performance of our algorithm on synthetic RCM images.

3.4.2 Accuracy on real RCM images

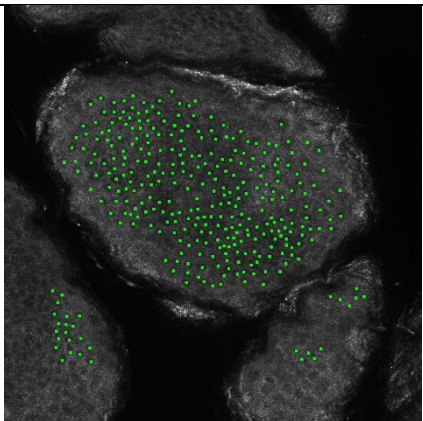
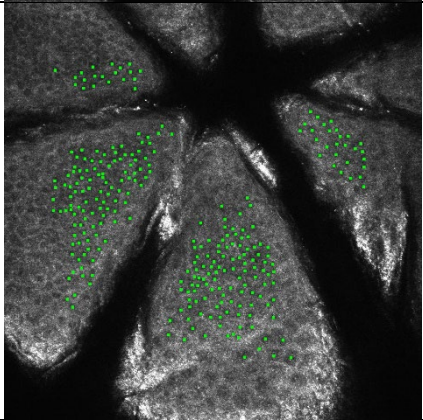
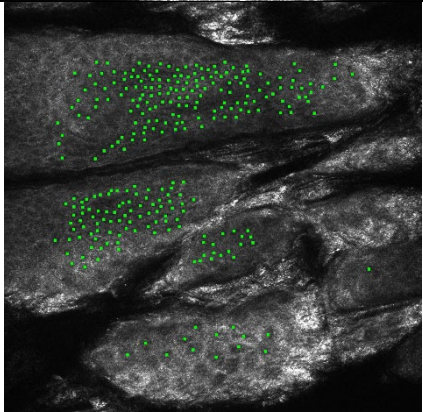
We evaluated the performance of the FIAP on 9 RCM images (images acquired on both check and volar forearm for participants ages 5 months to 35 years old during 3 clinical studies), where the ground truth has been evaluated by Expert 1 (see Table 4), resulting in a precision of 63.9% ($\pm 11.3\%$), and a recall of 79.6% ($\pm 12.1\%$). Three (3) of these images represent the spinous layer (precision = 48.9% ($\pm 2.9\%$), recall = 79.6% ($\pm 6.9\%$)). And the remaining 6 represent the granular layer (precision = 67.7% ($\pm 7.9\%$), recall = 79.4% ($\pm 14.2\%$)).

While the overall results were satisfactory and greater than the minimum accuracy threshold previously defined, we observed lower precision than recall for RCM images of both SG and SS, and significantly lower precision for keratinocyte detection in the SS and slightly lower recall for keratinocyte detection in the SG. However, the F1-score $= \frac{Precision * Recall}{Precision + Recall}$, which combines both accuracy metrics, was overall higher for the segmentation of RCM images of the SG. The lower the precision, the higher the number of false positives, *i.e.*, the number of invented cells. The lower detection precision in SS may be explained by the smaller cell size in this layer and therefore an increased difficulty in the segmentation with a tendency for over-segmentation.

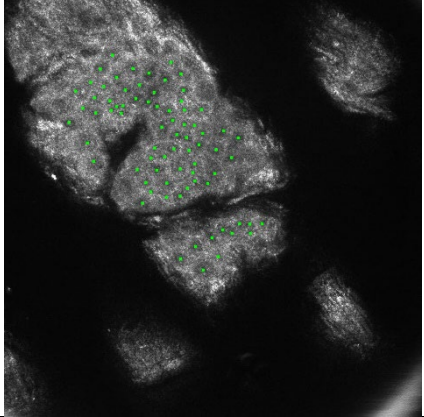
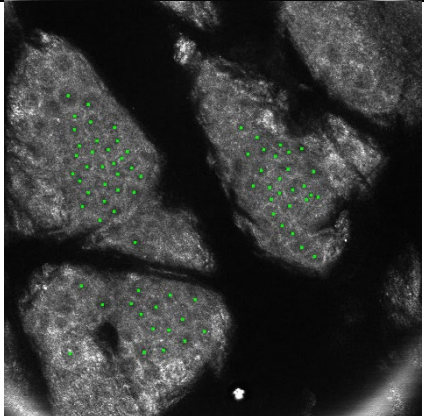
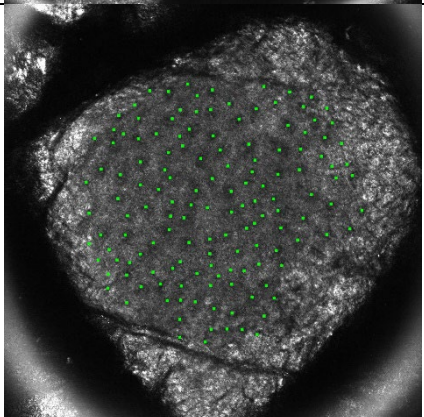
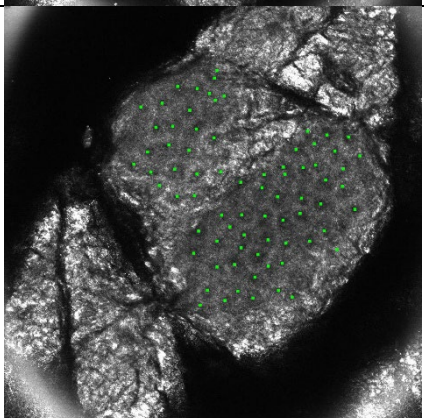
We noticed that all median measures have a relatively important standard deviation reflecting an important variation in image quality and noisiness.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

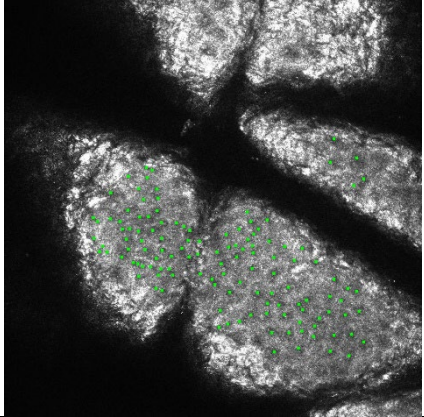
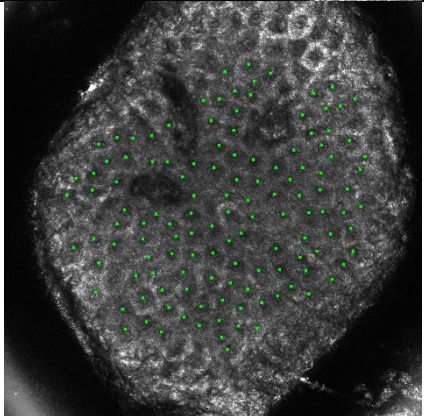
Table 4 Cell detection accuracy on 6 *stratum granulosum* and 3 *stratum spinosum* RCM images for Expert 1. Ground truth in the form of cell centers is represented in green on each image. Median data shown as median (± 1 standard deviation). RCM, Reflectance confocal microscopy.

Image	Layer	Precision (%)	Recall (%)	F1-score (%)
	SS	46.2	89.4	61.0
	SS	52.0	76.1	61.8
	SS	48.9	79.6	60.6

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Image	Layer	Precision (%)	Recall (%)	F1-score (%)
	SG	63.9	60.5	62.2
	SG	56.4	77.0	65.1
	SG	71.3	81.8	76.2
	SG	71.6	90.7	80.0

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Image	Layer	Precision (%)	Recall (%)	F1-score (%)
	SG	78.9	56.3	65.7
	SG	64.1	87.4	74.0
Median accuracy for images of the SG		67.7 (± 7.9)	79.4 (± 14.2)	69.9 (± 7.2)
Median accuracy for images of the SS		48.9 (± 2.9)	79.6 (± 6.9)	61.0 (± 0.6)
Overall median accuracy		63.9 (± 11.3)	79.6 (± 12.1)	65.1 (± 7.4)

The 6 RCM images of the SG were manually graded by a second expert (Expert 2) with less experience to assess the impact of training on manual segmentation quality and highlight the consequent inter-expert variability.

When compared to the first expert, the cell detection approach resulted in a precision of 71.4% ($\pm 7.4\%$) and a recall of 83.3% ($\pm 12.7\%$). When compared to the second expert, the cell detection approach resulted in a precision of 71.4% ($\pm 6.7\%$) and a recall of 61.7% ($\pm 15.7\%$) (see Table 5). When the two expert gradings were compared to each other, we obtained a precision of 59.6% ($\pm 5.6\%$) and recall of 40.5% ($\pm 11.9\%$). This result demonstrated high inter- (and intra) expert variability especially with regard to the recall which decreases with the

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

increase of false negatives, *i.e.*, undetected cells, showing that Expert 1 (more experienced) is more reliable.

The variability in expert opinions can significantly affect the accuracy of the ground truth we rely on to evaluate the effectiveness of our approach. To mitigate this potential issue, it would have been advantageous to establish a consensus by averaging results from multiple experts as a more reliable ground truth. Unfortunately, due to the limited number of experts available, this approach was not feasible for our study.

The comparison between the two experts confirms the importance of and need for an automated, unbiased, reproducible, keratinocytes segmentation method on RCM images.

Table 5 Cell detection accuracy on 6 RCM images of the *stratum granulosum* for two different experts. Data shown as median (\pm 1 standard deviation). RCM, Reflectance confocal microscopy.

	Precision (%)	Recall (%)	F1-score (%)
Detections vs. Expert 1	71.4 (\pm 7.4)	83.3 (\pm 12.7)	75.6 (\pm 4.9)
Detections vs. Expert 2	71.4 (\pm 6.7)	61.7 (\pm 15.7)	67.3 (\pm 9.7)
Expert 1 vs. Expert 2	59.6 (\pm 5.6)	40.5 (\pm 11.9)	48.0 (\pm 8.9)

While our approach demonstrates reasonable performance when applied to RCM images of the granular and spinous layers, its accuracy can be compromised when cells from different epidermal layers are present within the same image. This complexity makes parameterization of the various steps in the FIAP challenging. Our approach follows a multi-step process, with each step involving multiple parameters that influence cell detection and its accuracy, but with one set of parameters per epidermal layer (see Table 6). Furthermore, the presence of noise and non-uniformity in the images significantly impacts the performance of the pipeline. Although steps such as median filtering, local normalization, and ROI determination help mitigate the impact of noise on the results, they do not completely eliminate it.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Table 6 Parameters for each step on the FIAP for RCM images of the SG and SS.
FIAP, Full image analysis pipeline; RCM, Reflectance confocal microscopy; SG, *stratum granulosum*; SS, *stratum spinosum*.

FIAP step		Parameters value for RCM images of the SG	Parameters value for RCM images of the SS
Identification of the ROI	Snake: Borders between tissue and background detection	Image: Grayscale image Number of iterations: 500 Initialization level set: distance transform (mask size = 5, L2 distance) Smoothing: 5 Threshold: 50 Balloon: -1	
	SVM: non-informative areas detection and exclusion	Trained on the same GLCM features.	
	Bright areas detection and exclusion	Gaussian blur: kernel = (11,11), Binary thresholding: 200 Erosion: 2 iterations Dilation: 4 iterations	
Identification of individual cells	Median denoising	Size = 3 Mode: reflect	
	Sato filter	σ : [8,10], step 1 Black ridges: False Mode: Constant	Sigma: [5,6], step 1 Black ridges: False Mode: Constant
	Median denoising	Size = 7 Mode: reflect	
	Gabor filter	Kernel size: 41 σ : 5 λ : 50 γ : 10 θ : [0, π], step $\pi/64$	Kernel size: 41 σ : 5 λ : 25 γ : 10 θ : [0, π], step $\pi/64$
	Contrast Limited Adaptive Histogram Equalization	Clipping limit: 2 Tile grid (kernel) size: 10x10	Clipping limit: 2 Tile grid (kernel) size: 150x150
	Adaptive thresholding	Maximum value: 1 Adaptive method: Gaussian Threshold type: Binary Block size: 171 Constant: 10	
	Connected components analysis	Limit small objects size: 500	
	Connected components analysis on inverse of binary image	Limit small objects size: 15	
	Skeletonization & spurring	Number of iterations: 7	
Post-processing	Cleaning of skeleton	Opening with a ellipsoid kernel of size 31x31	
	Contour detection	Contour-retrieval mode: RETR_TREE	

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

FIAP step		Parameters value for RCM images of the SG	Parameters value for RCM images of the SS
		Contour-approximation method: CHAIN APPROX SIMPLE	
	Long contours removal at the borders with background	For contours at the border of the ROI, if eccentricity>0.85, remove the contour	For contours at the border of the ROI, if eccentricity>0.95, remove the contour
	Big contours selection (to be reviewed)	If contour area > median contour area + 2 standard deviations, review the contour in the following steps. Else it's a contour of the correct size.	
	Median filtering	Size = 3 Mode: reflect	Size = 5 Mode: reflect
	Gabor filter	Kernel size: 41 σ : 5 λ : 50 γ : 10 θ : $[0, \pi]$, step $\pi/64$	Kernel size: 41 σ : 5 λ : 25 γ : 10 θ : $[0, \pi]$, step $\pi/64$
	Local normalization	σ : 5	
	Sato filter on too big contours	σ : $[1,5]$, step 1 Black ridges: False Mode: Constant	
	Contrast Limited Adaptive Histogram Equalization	Clipping limit: 2 Tile grid (kernel) size: 171x171	
	Thresholding	Otsu	
	Connected components to remove small objects	Minimum objects size: 100	Minimum objects size: 50
	Skeletonization	(No manual parameters)	
	Contour detection	Contour-retrieval mode: RETR_TREE Contour-approximation method: CHAIN APPROX SIMPLE	
	Small contours merging	On the newly detected contours, while contour area<110 pixels, merge it to its neighbors.	On the newly detected contours, while contour area<20 pixels, merge it to its neighbors.
	Ellipse fitting when no new contours are detected	(Not in the pipeline)	Major axis length = Median Major axis length of the 10 closest correct contours

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

FIAP step		Parameters value for RCM images of the SG	Parameters value for RCM images of the SS
			Major axis length = Median Major axis length of the 10 closest correct contours Distance is calculated between contours centers.
	Cell centers detection	(No manual parameters)	

The FIAP processes a 1000x1000 pixels image in less than 10 minutes using 8 cores and 16 GiB of RAM. This time may vary depending on the size of the ROI and the level of noise in the image. This processing time should be compared to the 20-40 minutes required for manual annotation by an expert. This decrease in computational time significantly reduces the effort required to identify keratinocytes manually on RCM images. We will therefore move to applying it at large scale to over 5000 RCM images of the granular and spinous layer to derive quantitative parameters related to geometry and topology and use them to study epidermal architecture at the cellular level and examine the effects of age (0 to 80 years), body site location (cheek and volar forearm), and epidermal layer (SG and SS) on said parameters. To date attempts at automating cell identification on RCM images have been reported only on a limited number of images^{64,170}.

3.5 Application of the FIAP for the study of age and body site-dependent changes in epidermal structure

Studying the network organization of keratinocytes using the geometry and topology of the epidermis under a reflectance confocal microscope allows the researcher to quantitatively

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

analyze the structural patterns of the tissue and their evolution, their relationship to and impact on skin barrier function and therefore link to overall skin health⁹.

An essential element in understanding tissue development is the knowledge of cell positions and their features. Indeed, tissue arrangement plays a critical role in tissue development through cell communication and signaling, and spatial interaction between neighboring cells. Additionally, identifying abnormalities in cell positions and tissue architecture disarray (*e.g.*, uncontrolled growth and cell invasion) can provide insights into underlying disease mechanisms. Unfortunately, acquiring this information on RCM images is done manually, and therefore tedious, especially when done on a large number of images. Through the application of the FIAP to over 5000 RCM images, we were able to detect keratinocytes and extract the quantitative parameters of interest and thus study their dynamic evolution across different populations (age, body site, and epidermal layer) validating knowledge on skin maturation and ageing. To our knowledge, this is the first large-scale analysis of RCM images.

3.5.1 Statistical analysis

To represent each cell population depending on age group, epidermal layer, and body site, their median value will be used. Quantitative comparison was performed using an ANOVA on a fitted linear model. Additionally, we performed linear regression for each parameter per body site and epidermal layer to assess the relationship between age and the respective parameter, examining both the direction and strength of the relationship. Statistical significance was considered for $p\text{-value} < 0.05$. All statistical analysis were conducted using Python 3.6.7.

The FIAP was applied to all images resulting in the detection of keratinocytes present in every image and derivation of cell contours and centers which were used to calculate parameters relating to cell geometry such as cell area, perimeter, density, and relating to tissue topology such as the number of Delaunay nearest neighbors.

3.5.2 Cell geometry evolution with age and epidermal layer

Cell geometry (cell area, cell density, cell perimeter) gradually changed with age and with epidermal layer. Cells were significantly larger in the volar forearm compared to the cheek for both studied epidermal layers (see Fig. 16). The differences in cell geometry between children and adults were only significant for SG keratinocytes on the volar forearm, not on the cheek for neither SG nor SS, nor for spinous keratinocytes on the volar forearm.

We evaluate the gradual change in cell area with respect to age with a linear model for each epidermal layer (see Fig. 17). This change was only significant for granular keratinocytes on the volar forearm with $R^2 = 0.465$.

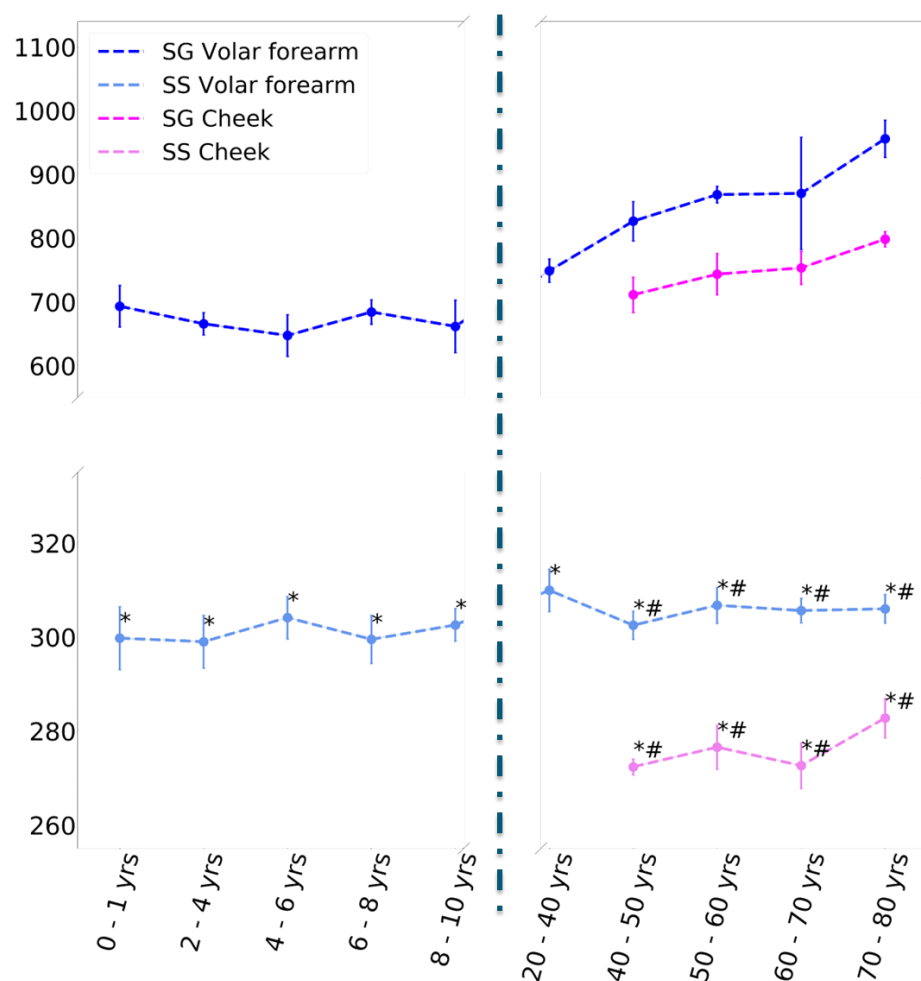


Fig. 16 The median cell area \pm standard error of mean per age group, epidermal layer and body site reflect the dynamic maturation and ageing of the epidermis. * Indicates that the median cell area is significantly different between the SG and the SS for a body

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

site and age group. # Indicates that the median cell area is significantly different between the cheek and the volar forearm per age group and epidermal layer. SG, *stratum granulosum*; SS, *stratum spinosum*.

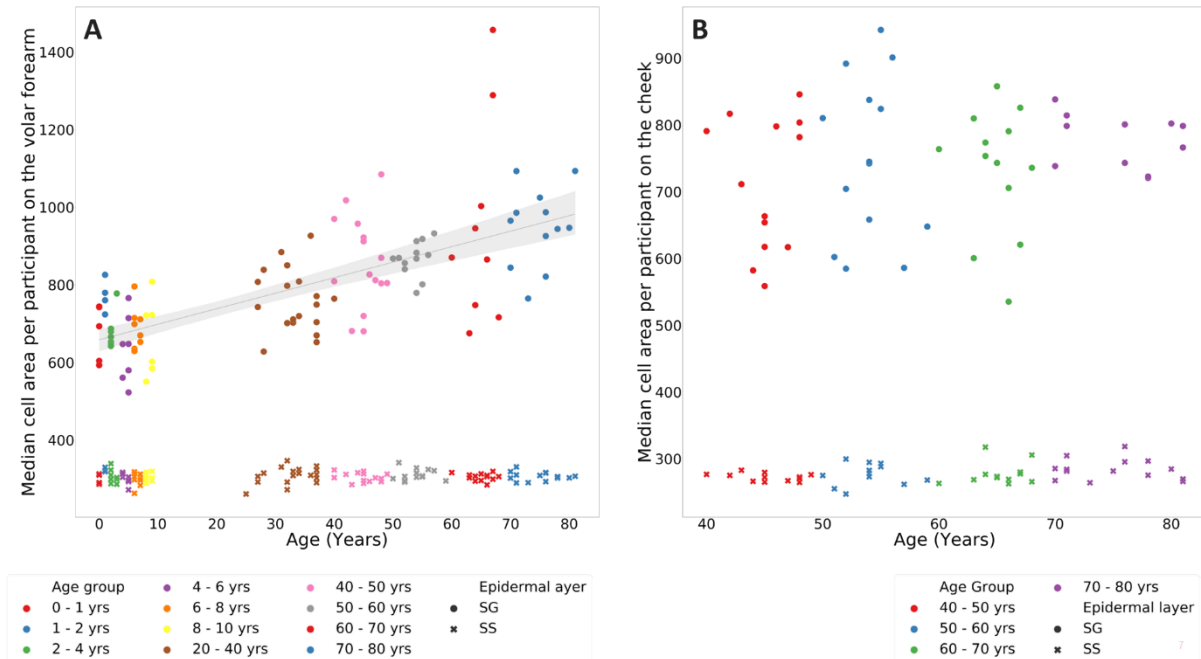


Fig. 17 Median cell area per participant on (A) the volar forearm (SG $R^2 = 0.465$; SS $R^2 = 0.00693$) and (B) the cheek (SG $R^2 = 0.0324$; SS $R^2 = 0.0708$) colored by age group. We fit a linear regression for each epidermal layer and body site but plot it only when significant. SG, *stratum granulosum*; SS, *stratum spinosum*.

With the application of the hybrid deep learning model for epidermal layer classification¹⁶⁵ on each RCM stack, we were able to calculate the thickness of the SC and of the supra-papillary epidermis (SPE) as the difference in depth between the uppermost and lowermost optical sections containing the desired observable features. As multiple stacks were captured per participant and per body site, we summarized the SC and SPE thicknesses per stack into a median value per study participant and body site.

Both SC and SPE thicknesses increase significantly during childhood with respect to age ($R^2 = 0.187$ for SC and $R^2 = 0.279$ for SPE) but not in adults (see Fig. 18).

These findings suggest that cell turnover is faster on the face than on the arm and is faster in children than in adults^{171,172} as cells reach the surface faster to be shed due to thinner epidermal

layers, and is associated with a higher keratinocytes proliferation rate¹⁷³, leading to a thickening of epidermal layers and an increase in cell size. The measured quantitative geometrical values confirm the current physiological knowledge about epidermal maturation. Indeed, even though epidermal structure and barrier function is competent at birth and during childhood, the epidermis is more susceptible to outside-in aggressors (penetration of noxious substances^{172,174–176} due to lower glyph density) and inside-out issues (water evaporation leading to tissue desiccation¹⁷⁷ due to higher TEWL, higher conductance, and lower natural moisturizing factors in infants^{176,178}). These functional differences are correlated with the structural geometrical changes previously described.

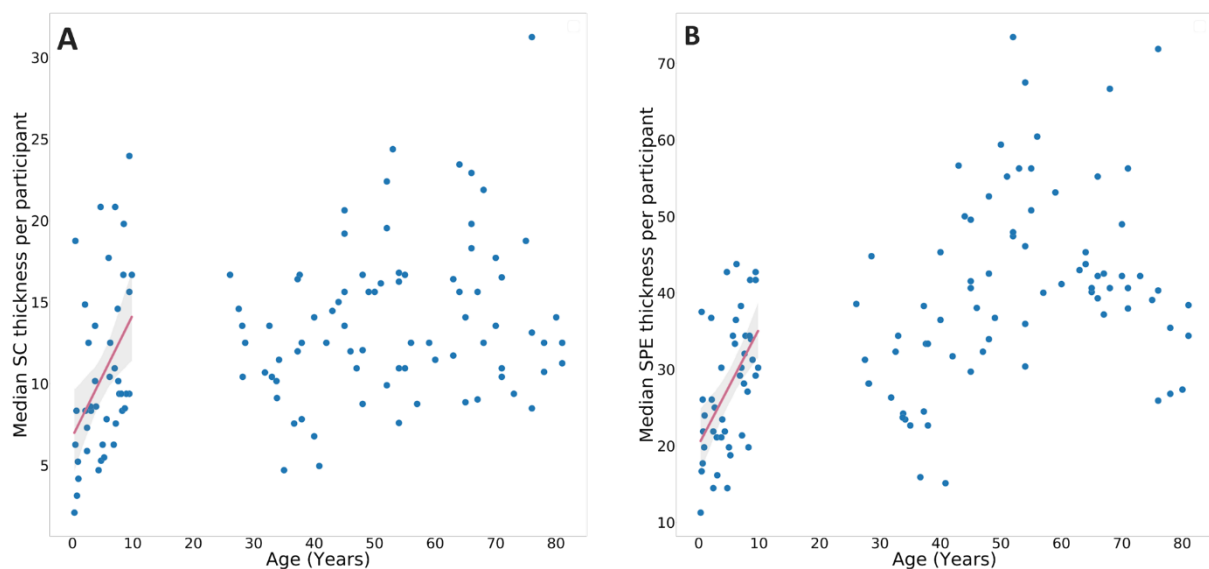
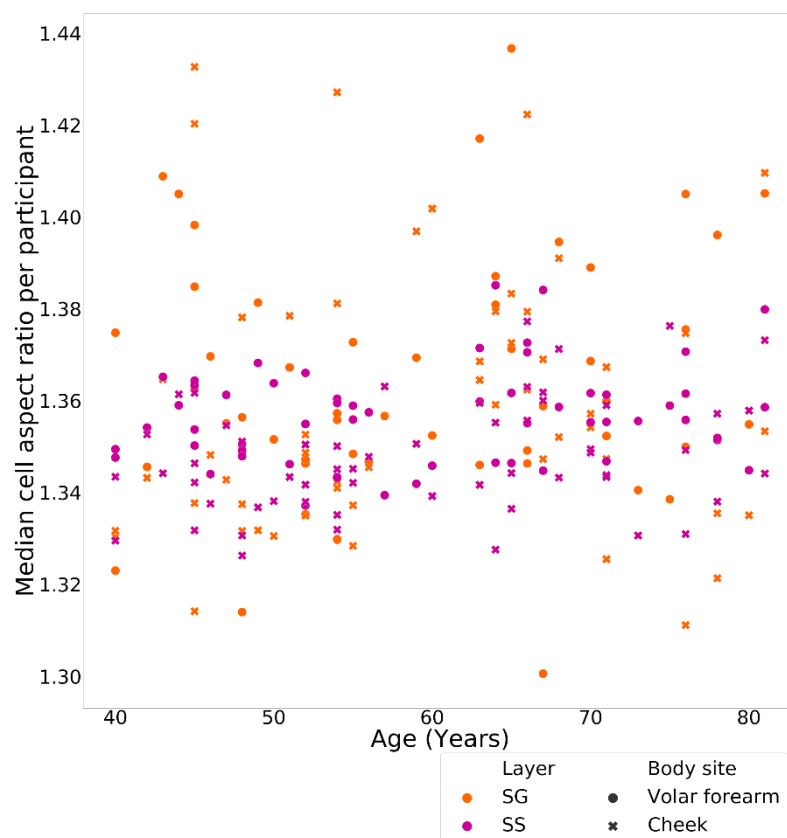


Fig. 18 (A) Median SC thickness per participant (children $R^2 = 0.187$), **(B)** Median SPE thickness per participant (children $R^2 = 0.279$). A linear correlation with age was not significant in the adult group for either the SC or the SPE thickness. SC, *stratum corneum*; SPE, supra-papillary epidermis.

Measurement of these geometrical parameters and understanding of epidermal architecture provide valuable insights into skin barrier function and skin health, and aid in the understanding of various skin condition. Indeed, epidermal layers thickness are of particular interest as they play a role in the determination and normalization of topical drugs permeation profile in

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

dermato-pharmacokinetic studies¹⁷⁹ for the study of their bio-availability and bio-equivalence between formulas. They are also an essential consideration when developing cosmetic products to determine optimal formulation to ensure products efficacy and safety. Additionally, the epidermis is often represented as a brick and mortar wall¹⁸⁰, with the keratinocytes considered as the brick, they are therefore essential in maintaining a healthy skin barrier function, highlighting the importance of the study of their evolution and differentiation from one layer to another and how they may alter the skin barrier function, resulting in various skin alterations¹⁸¹. While the SG and SS do not directly contribute to the main barrier function of the skin, the cell size of keratinocytes in these layers can indirectly influence the formation of a robust and effective skin barrier through the processes of keratinization, lipid production, tight junction, and corneocyte formation. These factors collectively contribute to the overall function and integrity of the skin barrier.



A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Fig. 19 Median cell aspect ratio per participant. SG, *stratum granulosum*; SS, *stratum spinosum*.

Surprisingly, cell aspect ratio ($\frac{\text{Major axis length}}{\text{Minor axis length}}$) does not seem to be impacted by the previously described changes. Indeed, through the differentiation of keratinocytes, as they move towards the surface through the spinous and granular layer, and the changes this may entail, their overall aspect and relative dimensions do not vary (see Fig. 19).

3.5.3 Cell topology evolution with age and epidermal layer

We initiated a Delaunay triangulation from the identified cell centers on each image and used it to calculate the average distribution of Delaunay nearest neighbors per age group, epidermal layer, and body site (see Fig. 20). A Delaunay triangulation consists of a net of triangles that guarantees that the circumcircle of each triangle contains only the vertices of said triangle¹⁸². The Delaunay triangulation of a set of points is dual to the Voronoi diagram initiated with these points as seeds. Indeed, the circumcenters of the Delaunay triangles form the vertices of the Voronoi diagram. Both concepts have been used to study cellular sociology^{9,166,183} to explore the relationship between cells in both healthy and diseased tissue.

While geometrical parameters change with age, body site, and epidermal layer, structural organization does not. It seems to be preserved through epidermal maturation and ageing in healthy skin, and through the differentiation process of keratinocytes.

We also calculated the Fisher-Pearson coefficient of skewness¹⁸⁴ of the probability distribution for each age group, epidermal layer, and body site, as a measure of distribution asymmetry (see Fig. 22 and Table 7). The obtained values are compared to those of two published models of healthy and cancerous tissue (see Fig. 21). These models are based on game theory¹⁸⁵ to model interactions between cells using different strategies, each one associated with a genotypic

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

variant linked to a different fitness and cell division speed, specifically an evolutionary game known as the prisoner's dilemma where a cooperator pays a cost for the distribution of benefits to its partners/neighbors, while a defector (cheater) does not pay any cost and does not yield any benefit to its partners. The balance between benefits and costs impacts the fitness of the studied population, allowing a spread of defectors in a population of cooperators as they have less costs and can replicate faster. Here, the cooperators model represents healthy epithelium, while the defectors model represents cancerous tissue, where cells have lost some of the controls from other cells and their environment leading to a different behavior.

The obtained skewness values (see Table 7) are closer to that of the cooperators model with a skewness of 0.40, *i.e.*, the data distribution is symmetrical, and much lower than that of the defectors model with a skewness of 0.98, *i.e.*, greatly skewed data distribution. There are no statistically significant differences between skewness values of the SG and SS distributions on both studied body sites.

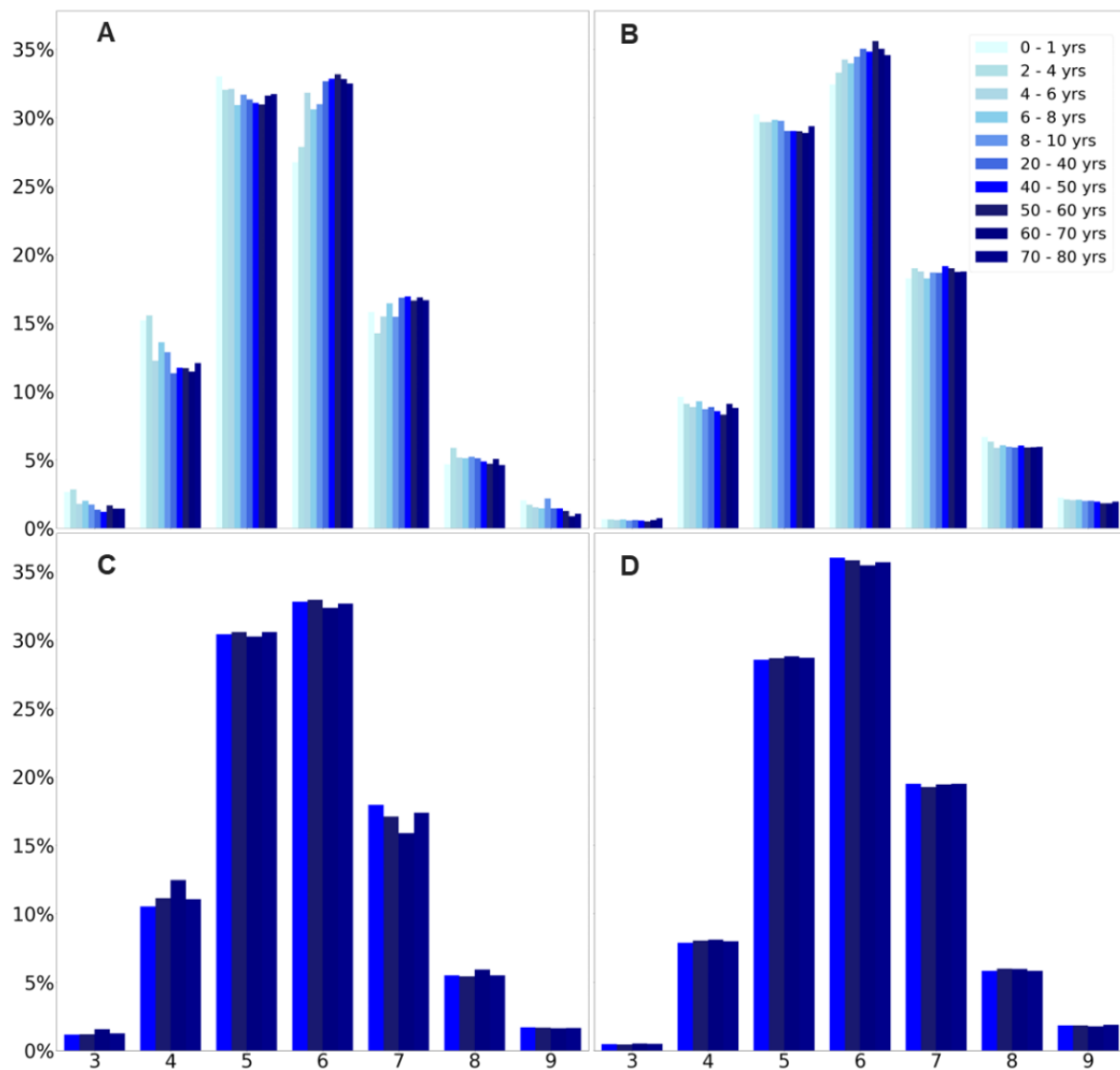


Fig. 20 Average probability distribution of the number of Delaunay nearest neighbors per cell for **(A)** SG on the volar forearm, **(B)** SS on the volar forearm, **(C)** SG on the cheek, and **(D)** SS on the cheek per age group. SG, *stratum granulosum*; SS, *stratum spinosum*.

We have shown that the epidermis, specifically the SG and SS, constantly undergoes changes throughout its maturation, ageing, and differentiation. The obtained results validate previously published work on these topics⁹ and extend it by studying a larger age range across the population using thousands of images. This quantitative study of epidermal cells spatial organization was made possible by the use of the FIAP, developed in this thesis, for the automated detection of keratinocytes on RCM images of the granular and spinous layers.

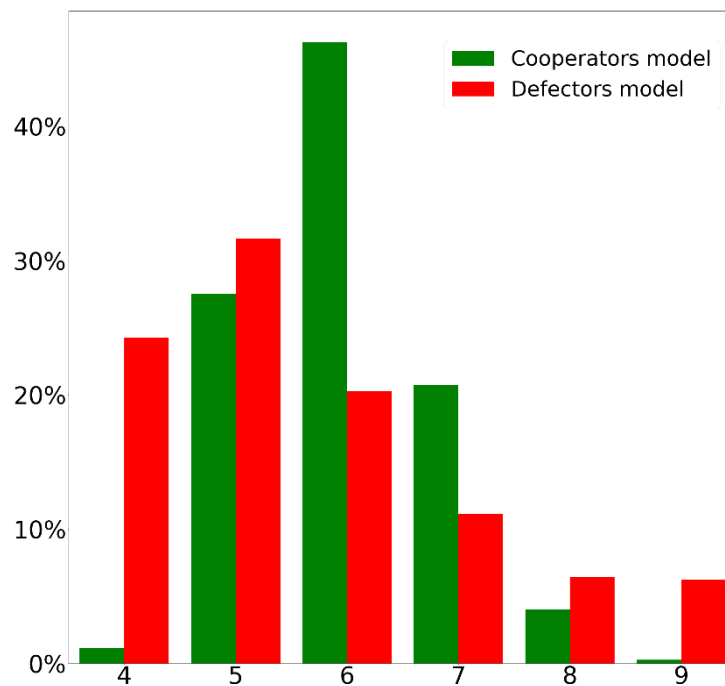


Fig. 21 Probability distribution of the number of Delaunay nearest neighbors per cell for the cooperators model (in green), and defectors model (in red).

By automating a previously manual and laborious step, we have significantly enhanced our capacity to analyze RCM images. This advancement has enabled us to examine a greater number of images than ever before, solidifying our understanding of the structural maturation of the epidermis. Moreover, it has facilitated the expansion of our study across a broader age range, encompassing individuals from birth to 80 years old, thus allowing a more comprehensive analysis of the geometrical and topological properties of the epidermis across ages and body sites, in contrast to previous studies that primarily focused solely on comparing infant skin to adult skin ^{171,172,174–176,178}.

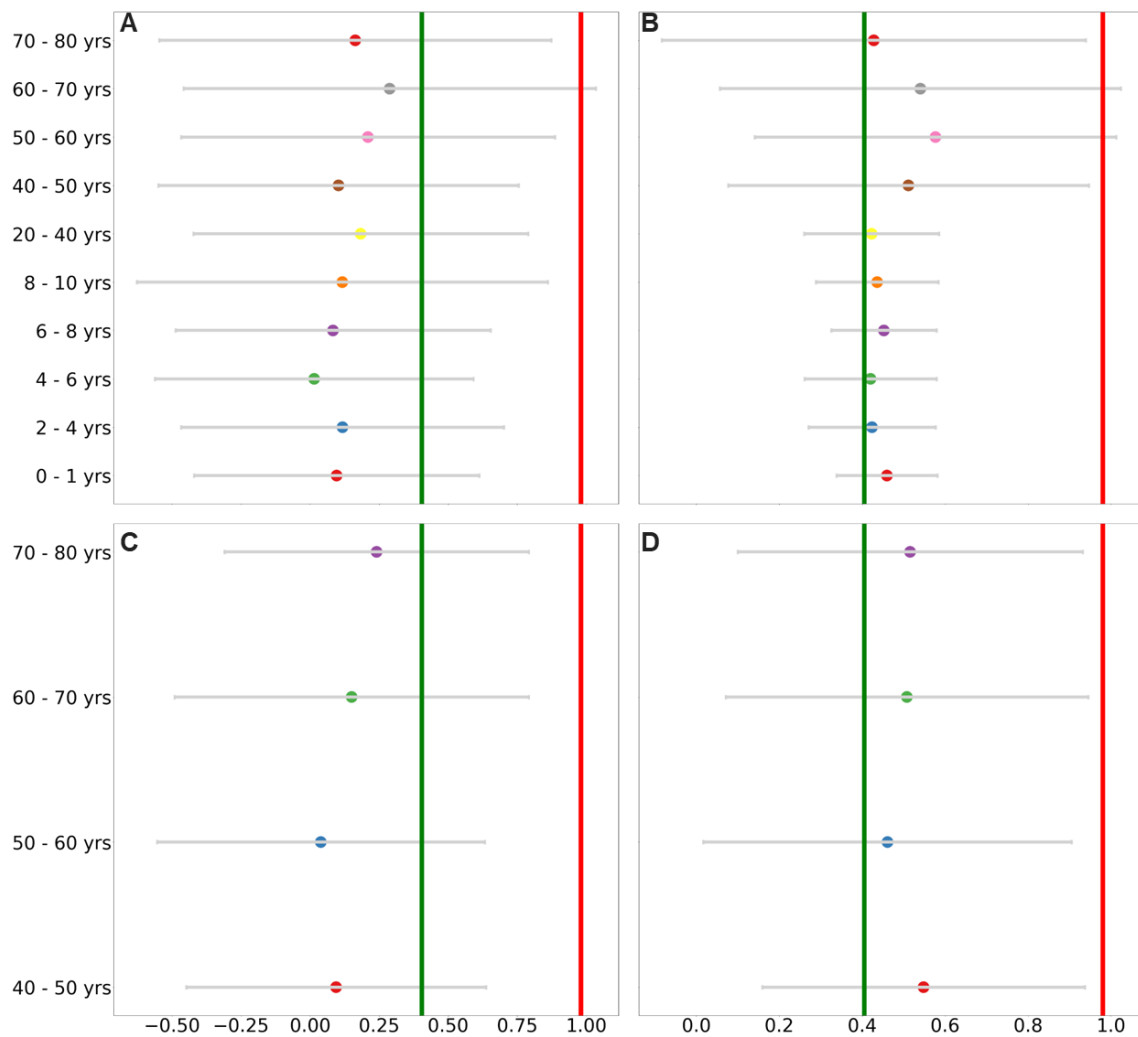


Fig. 22 Distribution of skewness per age group for (A) SG on the volar forearm, (B) SS on the volar forearm, (C) SG on the cheek, (D) SS on the cheek. In red, skewness of the probability distribution of the Delaunay nearest neighbors for the defectors model. In green, skewness of the probability distribution of the Delaunay nearest neighbors for the cooperators model. SG, *stratum granulosum*; SS, *stratum spinosum*.

The findings of the application of the FIAP to thousands of RCM images align with previous research on the maturation of the epidermis. It validates that the increase in cell size observed in adults compared to children is likely attributed to the deceleration of cell turnover with age¹⁸⁶. Moreover, the cells in the granular layer were found to be larger than those in the spinous layer, indicating a progressive increase in size during maturation. This can be attributed to the prolonged residence time of cells in the epidermis due to reduced cell turnover¹⁸⁷, allowing for

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

more extensive maturation and growth before eventual desquamation at the skin surface. Additionally, the quantitative analysis revealed that the *enface* surface area of cells is bigger on the volar forearm compared to the cheek which reflects the higher cell turnover rate on the face¹⁸⁷ which also decreases faster with age.

Although the spatial organization of healthy skin cells undergoes dynamic changes clear in the evolution in cell sizes with age, their topology (spatial arrangement) remains consistent. We observed that the skewness of the probability distribution of Delaunay nearest neighbors is greater in the spinous layer compared to the granular layer. Interestingly, this skewness is closer to that observed in the defectors model, suggesting a resemblance between the spatial patterns of cells in the spinous layer and the behavior of defectors in the model. While the observed differences may not reach statistical significance, we could hypothesize that the proximity of the spinous layer to the basal layer, where keratinocytes replicate, could influence the probability distribution of Delaunay nearest neighbors. This could potentially reflect the proliferation process occurring in the basal layer, with the spinous layer displaying a distribution closer to the defectors model where cancerous cells adopt a cheating competitive strategy, leading to abnormal and excessive proliferation. Additional analysis of both the spinous and basal layers is however required to validate this hypothesis.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Table 7 Average Fisher-Pearson coefficient of skewness of the probability distribution of Delaunay nearest neighbors per age group. SG, *stratum granulosum*; SS, *stratum spinosum*.

	SG volar forearm	SS volar forearm	SG cheek	SS cheek	Cooperators	Defectors
0 – 1 yr.	0.226	0.465	-	-	0.404	0.980
2 – 4 yrs.	0.229	0.429	-	-		
4 – 6 yrs.	0.135	0.444	-	-		
6 – 8 yrs.	0.0949	0.457	-	-		
8 – 10 yrs.	0.0354	0.448	-	-		
20 – 40 yrs.	0.244	0.434	-	-		
40 – 50 yrs.	0.282	0.574	0.229	0.607		
50 – 60 yrs.	0.332	0.592	0.131	0.493		
60 – 70 yrs.	0.427	0.593	0.228	0.566		
70 – 80 yrs.	0.412	0.465	0.320	0.551		
Median	0.135	0.444	0.229	0.559		

Although the size of granular and spinous keratinocytes changes with age, epidermal layer, and body site, their shape is preserved through cell maturation, ageing, and differentiation. Conducting a similar analysis on proliferating basal cells would be valuable since the examination of cellular aspect ratio has primarily focused on proliferative or cancerous tissues^{188–190}. In such studies, cellular aspect ratio has been identified as a fundamental parameter for normal cell division and function. Furthermore, it has been proposed as a potentially vital factor in maintaining cell geometry within proliferative tissues, as well as in shaping the spatial patterns of daughter cells during cell movement and differentiation.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Therefore, exploring the aspect ratio of proliferating basal cells can provide valuable insights into the mechanisms governing cell division, cell function, and the establishment of spatial patterns in tissues undergoing proliferation. By investigating this aspect, we can further enhance our understanding of cellular dynamics and their role in tissue development and maintenance.

This study only included Caucasians with Fitzpatrick skin phototypes from I to III. Nevertheless, the FIAP was also tested on two RCM images of the granular layer of subjects with heavily pigmented skin with similar accuracy compared to a manual ground truth obtained by Expert 1 (see Table S1). It would be valuable to extend this work to include more participants with different phenotypic and genetic background including more pigmented skin.

To summarize, our study successfully demonstrated the applicability of an automated approach for detecting keratinocytes in RCM images. This method enabled the extraction of geometrical and topological properties, facilitating comparisons across age groups, epidermal layers, and two body sites. By using the FIAP, we were able to overcome a common challenge in biomedical image analysis, namely the laborious nature of manual tasks and the potential for human interpretation biases. By extending the proposed method to RCM images of diseased skin, we could generate novel insights into various skin conditions and uncover valuable knowledge about the characteristics and behaviors of keratinocytes in different skin disorders. This, in turn, may contribute to advancing our understanding of skin diseases and potentially lead to improved diagnosis and therapeutic strategies. This last point was unfortunately not possible due to lack of access to RCM images of diseases skin.

3.6 Comparison to machine learning based approaches

The FIAP succeeded in detecting keratinocytes on RCM images of the granular and spinous layers and was successfully applied to thousands of RCM images, but the method can be

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

hindered by image noise, lack of contrast and non-uniformity, and the presence of cells from different epidermal layers on the same image making parametrization of the different steps complicated.

To address these challenges, we tested machine learning-based methods. This shifted our perspective from descriptive analysis, *i.e.*, using existing knowledge or prior information about the morphological characteristics of the studied membranes, to predictive analysis, *i.e.*, training a model to unveil underlying patterns within the images by minimizing disparities between the ground truth and predictions. The aim was to minimize the need for manual adjustments and decrease computational time.

However, a significant challenge in adopting this kind of approach in the case of RCM images, and biomedical imaging in general, is the limited availability of labeled images for training. To overcome this limitation, we used the previously described synthetic images for data augmentation. This approach allowed us to generate additional data to hopefully enhance the model learning capabilities and improve its predictive performance.

We tested two machine learning based approaches: the U-net algorithm¹⁹¹ and the Cell Cutter algorithm¹⁹². Images were split between training and testing with an 80-20 ratio and were the same for all U-net models.

The previously described U-net algorithm, made of a contracting and expanding path to capture contextual information while conserving spatial information, was tested in 6 different configurations, and their accuracy evaluated (see Table 8, Fig. 23). Images used in U-net were of size 256x256, resulting from the splitting of 1000x1000 full framed RCM images, to reduce training time.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

A common problem encountered when training machine learning models is the class imbalance issue between background and foreground. To counter this issue, we chose to use a ground truth in the form of membranes positions rather than cell centers position when training machine learning models. This ground truth was obtained by generating a Voronoi tessellation from manually obtained cell centers and will be used in the training of the model. However, accuracy will be calculated by comparing detected cells to cell centers positions as previously done in Chapter 3, in order to not introduce any bias to the calculated accuracy metrics.

The first tested U-net model was trained on a small dataset of 39 real RCM images of 4 participants, ages 20 to 35 years. The scarcity of manually labeled RCM images, where ground truth is known, is due to the time and expertise required to identify keratinocytes on said images. As a consequence, our initial attempt with this restricted dataset yielded a very low accuracy.

To address this challenge, we tested a second U-net configuration and augmented the dataset by incorporating synthetic images. This augmentation proved beneficial, enhancing the accuracy metrics for both real and synthetic images (see Table 8). However, the results were still very much unsatisfactory for real RCM images (see Fig. 23), especially the trade-off between recall and precision, visible in the low F1-score value. The very low recall reflects the high number of undetected cells, while the high precision indicates that the detected cells match the ground truth.

We then moved to testing a U-net model pretrained on the 2012 ImageNet Large Scale Visual Recognition Challenge dataset¹⁹³ with an efficientnetb3 backbone¹⁹⁴. This attempt assumes that a pretrained model will be closer to convergence and therefore more likely to yield accurate segmentation results with a smaller training set. We first used this pretrained U-net model without additional training (third tested U-net model), as an accuracy baseline and an evaluation of the impact of additional data on model performance, and then with supplementary real and

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

synthetic RCM images (fourth tested U-net model), and different loss functions to account for the class imbalance in RCM images, *i.e.*, there is much more background than there are cell membranes on RCM images. We tested out the Dice loss function¹⁹⁵ defined as follows:

$$Dice\ Loss\ (y, \hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1} \quad (10)$$

with $(y, \hat{p}) = (real\ value, predicted\ value)$. This loss function is derived from the Sørensen-Dice coefficient and measures the dissimilarity between the predicted segmentation mask and the ground truth mask by calculating the overlap between the two.

and the Focal loss function¹⁹⁶ defined as follows:

$$Focal\ Loss\ (p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (11)$$

with the estimated probability of class p_t defined as:

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (12)$$

with $(y, p) = (real, prediction)$, and α_t and γ weight parameters. The first addresses the class imbalance and the second controls the degree of emphasis on hard examples. The Focal loss addresses the issue of class imbalance in object detection problems by mitigating the dominant effect of the abundance of easy to classify background. This is done by giving more weight to “hard” misclassified examples while down-weighting “easy” correctly classified ones.

Our final tested U-net model assessed the performance of a combination of the two loss functions. We achieved similar accuracy values across all three configurations when evaluating synthetic images. For real RCM images, the best cell identification accuracy was obtained with a pre-trained U-net model augmented with both real and synthetic RCM images and using the combined Dice and Focal loss functions (see Table 8). However, it is important to note that

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

despite these advancements, the accuracy of all tested U-net approaches remained lower than that of the previously presented FIAP, particularly concerning real RCM images.

Table 8 Accuracy metrics for the tested U-net models.

Model	Image dataset	Synthetic RCM images			Real RCM images		
		Median testing precision	Median testing recall	Median testing F1-score	Median testing precision	Median testing recall	Median testing F1-score
U-net trained on real images only	Training: 39 images Testing: 10 images	0	0	0	—	—	—
U-net trained on real and synthetic images	Training: 280 images Testing: 70 images (6 of which are real RCM images)	0.923 ± 0.092	0.958 ± 0.093	0.930 ± 0.086	1.000	0.083 ± 0.108	0.154 ± 0.156
U-net pre-trained with no additional training	Pre-training on 2012 ILSVRC ImageNet dataset, with efficientnetb3 backbone	0.027 ± 0.020	0.091 ± 0.107	0.041 ± 0.031	0.017 ± 0.008	0.125 ± 0.067	0.031 ± 0.014
U-net pre-trained and augmented with real and synthetic images with Dice loss function	Pre-training on 2012 ILSVRC ImageNet dataset, with efficientnetb3 backbone Training: 203 synthetic	0.909 ± 0.167	0.913 ± 0.107	0.911 ± 0.147	0.520 ± 0.109	0.482 ± 0.234	0.516 ± 0.164

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

Model	Image dataset	Synthetic RCM images			Real RCM images		
		Median testing precision	Median testing recall	Median testing F1-score	Median testing precision	Median testing recall	Median testing F1-score
U-net pre-trained and augmented with real and synthetic images with Focal loss function	images, 43 real images Validation: 68 synthetic images, 13 real images Testing: 30 synthetic images, 5 real images	0.917 ± 0.154	0.923 ± 0.107	0.923 ± 0.136	0.500 ± 0.044	0.609 ± 0.238	0.550 ± 0.120
U-net pre-trained and augmented with real and synthetic images with Focal and Dice loss functions		0.909 ± 0.174	0.917 ± 0.100	0.909 ± 0.148	0.545 ± 0.159	0.571 ± 0.186	0.603 ± 0.154
FIAP	—	0.835 ± 0.067	0.925 ± 0.012	0.878 ± 0.021	0.720 ± 0.068	0.850 ± 0.110	0.779 ± 0.084

The second machine learning approach tested was the Cell Cutter algorithm, which is an unsupervised marker-controlled segmentation method. This algorithm does not rely on manually annotated data for training. Instead, it employs marker locations generated from either real or synthetic nuclei images. Local U-net algorithms are then applied to each patch surrounding an identified marker to model cell features and improve the accuracy of membrane segmentation. By combining marker-based localization and U-net-based modeling, Cell Cutter theoretically offers an effective approach for automated cell segmentation without the need for

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

manual annotations in the training process. By employing this localized patch approach, we transform a complex multi-cell segmentation problem into a simpler multi-single-cell segmentation task. This means that each patch, when properly defined by its marker, contains only one cell.

Consequently, instead of simultaneously segmenting multiple cells, we focus on accurately detecting a single cell within each patch. This approach is built on the assumption that nuclei exhibit simpler morphological characteristics (or unique staining in histology), making them easier to detect to generate the markers. By leveraging this assumption, we aim to alleviate the under-segmentation bias often encountered in images with densely populated neighboring cells where boundaries between cells can be difficult to delineate accurately¹⁹⁷, as long as nuclei are correctly detected. Since we do not have images of the nuclei matching our RCM images (as with a DAPI image), we built synthetic marker images by applying the first two steps of our FIAP, *i.e.*, ROI and individual cells identification, thus using Cell Cutter as a replacement to our post-processing step. This was an attempt at striking a balance between a descriptive and a predictive logic that resulted in a hybrid approach combining the two aforementioned methods. However, when applying the Cell Cutter algorithm to our real RCM images, the results were unsatisfactory. The best achieved results for precision and recall were 71% and 56% respectively, falling below the acceptable thresholds. This indicates that the algorithm encounters challenges in properly distinguishing and segmenting cells, leading to lower accuracy values compared to the previously described FIAP.

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

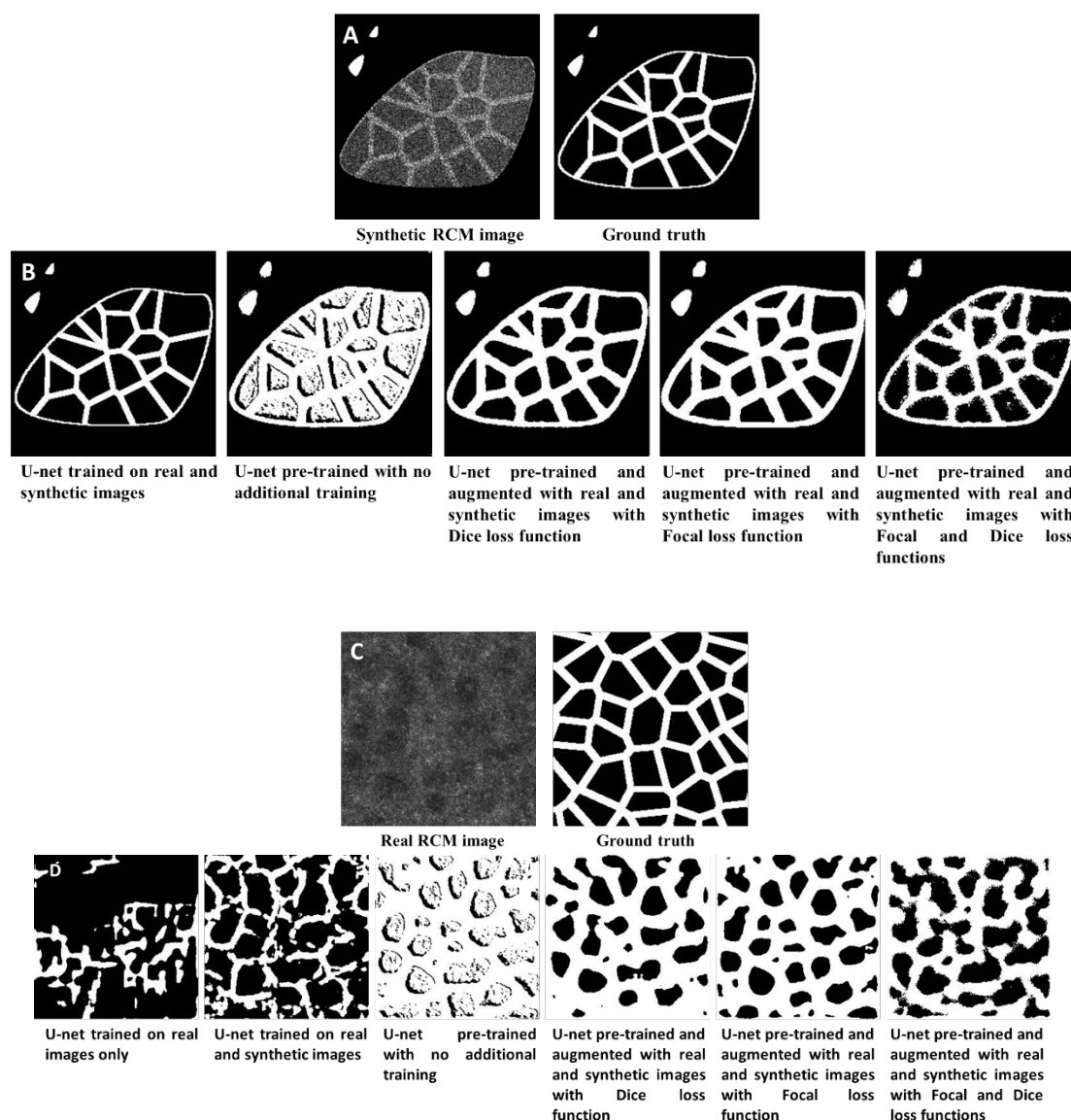


Fig. 23 (A) Synthetic RCM images and its segmentation ground truth. (B) Tested U-net architectures segmentations on a synthetic RCM image. (C) Real RCM images and its segmentation ground truth. (D) Tested U-net architectures segmentations on a real RCM image. RCM, Reflectance confocal microscopy.

3.7 Discussion & Conclusion

Our study successfully demonstrated the feasibility of automated keratinocyte detection on RCM images of the granular and spinous layers using a morphology-based approach. This is an important step towards streamlining the quantitative analysis of these images to further our

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

understanding of healthy and diseased epidermis in both clinical application and basic research settings.

However, we should acknowledge that the accuracy and computational efficiency of our method can be influenced by various factors: required epidermal layer-dependent parametrization, and image noise and non-uniformity. To address these challenges, we explored machine learning-based methods as an alternative which unfortunately yielded lower accuracy values compared to the FIAP (see Tables 8 and 9, Fig. 23). This can be attributed to several factors, including the limited size of the training set comprising real images (ranging from 39 to 43, depending on the tested model as shown in Table 8) and the inherent differences between these images and the synthetic images used for data augmentation. This discrepancy between real and synthetic images may have affected the generalizability of the models, leading to lower accuracies. Furthermore, the accuracy of these approaches could have been impacted by the method used to obtain the ground truth on real RCM images. Indeed, the ground truth was generated using Voronoi tessellation using manually detected cell centers as seeds which may lead to an imperfect alignment between the Voronoi diagram vertices and the actual positions of the cell membranes. The rationale behind using artificially created membrane ground truth from manually annotated cell centers was to mitigate the issue of class imbalance in our images by shifting the problem from cell center detection to membrane segmentation, as there are typically more background pixels than cell centers and cell membranes. Additionally, subjectivity in the manual segmentation serving as ground truth, could have influenced the accuracy values. This variation in segmentation results was evident in the differences between the two experts (see Table 5). These factors collectively contributed to the observed differences in segmentation accuracies between real and synthetic RCM images in our study (see Table 8).

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using filtering and cell morphological features

A prospective solution to the limited labelled data and thus the low accuracy of machine learning based approaches could be using multi-task learning to perform multiple related tasks in parallel, with each task improving the others performance by constraining each other's solution space and thus improving overall accuracy.

Table 9 Comparison between the FIAP, Cell Cutter algorithm, and the U-net architecture. FIAP, Full image analysis pipeline.

	Advantages	Limitations
Manual cell identification	<ul style="list-style-type: none"> Fully explainable. 	<ul style="list-style-type: none"> Time consuming. Subjective.
Full Image Analysis Pipeline	<ul style="list-style-type: none"> Explainable: Based on knowledge of the morphological properties of the studied structures. Good accuracy against manual segmentation by expert graders. Satisfactory trade-off between recall and precision. 	<ul style="list-style-type: none"> Presence of multiple layers. Manual parametrization.
U-net	<ul style="list-style-type: none"> Based on prediction and discovering patterns in the image. Shorter computational time (excluding training time) Poor trade-off between precision and recall. 	<ul style="list-style-type: none"> Important training set size required for good results leading to poor accuracy on real RCM images.
Cell Cutter	<ul style="list-style-type: none"> Based on prediction and discovering patterns in the image. Shorter computational time (excluding training time) Multi-single cell segmentation instead of multi-cell segmentation. 	<ul style="list-style-type: none"> Important training set size required for good results leading to poor accuracy on real RCM images. Prior knowledge required: marker locations.

The proposed FIAP is now limited to RCM images of the SG and SS and was used in an extensive study of skin maturation and ageing across ages, epidermal layers, and body sites of lightly pigmented skin. It would be biologically interesting to extend its use to the analysis of

A traditional approach: Automatic granular and spinous epidermal cell identification and analysis on *in vivo* reflectance confocal microscopy images using filtering and cell morphological features

RCM images of the basal layer where cell replication occurs, and several skin diseases emerge. Unfortunately, this would be challenging using RCM, whether done manually or automatically. Indeed, it would be complicated for an expert to establish a ground truth on images of the SB because of poor image quality and severe drop in signal-to-noise ratio.

To the best of our knowledge, there has been limited research published on the automated detection of keratinocytes on *in vivo* RCM images. One paper that explored this area utilized a rotationally symmetric error function reflectance profile to model the shape of keratinocytes⁶³. However, this method employed fixed parameters for both the granular and spinous cells, and its accuracy was statistically validated based on cell density. In contrast, the proposed FIAP was validated against a manually obtained ground truth. This, we believe, makes our proposed method more accurate.

Despite its limitations, the approach gave satisfactory results in the detection of keratinocytes on RCM images of the SG and SS, and the normalization steps helped achieve a robust parametrization of the approach for each epidermal layer. Classical machine learning approaches failed to give satisfactory results, but more advanced deep learning methods based on multi-task learning could give more accurate results in keratinocytes detection on RCM images of the SG and SS.

Chapter 4 DermoGAN: An unsupervised multi-task deep-learning architecture for cell segmentation

The examination of RCM images provides valuable insights into the geometrical and topological characteristics of the epidermis. Typically, RCM stack analysis is carried out manually, resulting in qualitative observations. However, manual analysis is a time-consuming, labor-intensive process prone to inter-expert variability. Consequently, there is a need for automated techniques that can quantitatively analyze RCM images and identify the positions of keratinocytes within them.

To address this need, we introduced the FIAP, a method that relies on image filtering and the enhancement of membrane morphological features for keratinocyte detection in RCM images. We developed two distinct sets of parameters, one for the SG and another for the SS. These sets of parameters were determined using images from different clinical studies involving both children and adults and acquired on both the volar forearm and the cheek. Our results demonstrated that the FIAP successfully identifies keratinocytes on RCM images of both the granular and spinous layers. However, the manual parameterization it requires and the presence of multiple layers within one RCM image can impede its performance.

To overcome these limitations, we explored the application of classical machine learning models, such as U-net and CellCutter. Unfortunately, these attempts were unsuccessful due to the substantial amount of data required to achieve satisfactory performance and issues related to class imbalance between the foreground and background. Steps taken to circumvent these limitations were not enough to achieve a better performance than the FIAP. More advanced deep learning method may help achieve a good accuracy in keratinocytes identification on RCM images.

Accurate automated cell identification on biomedical images with deep learning has been a growing research topic in computer vision but is hindered by the lack of labelled data on account of cost, time, and domain-specific skills. Unsupervised learning bypasses the labeled data scarcity problem by tapping into the potential of unlabeled data. One of the main developments in unsupervised learning research of the recent years are cycle-GANs¹⁹⁸ for unpaired image-to-image translation, and are classically used for synthetic images generation and data augmentation^{199–202}.

We propose a top-down, structure aware, multi-task cycle-GAN architecture, named DermoGAN, to automatically detect keratinocytes on RCM images. The multi-task model performs two parallel cycle-GANs to denoise RCM images while highlighting membranes positions, and provides an incomplete cell identification, which is then refined and completed by a post-processing based on star-convex shape segmentation, *i.e.*, segmenting detected contours such that they are all star-convex polygons. Star-convex polygons are shapes that contain a point called center such that, for any point within the polygon there exists a line segment connecting the point to the center within the polygon. The proposed architecture is fully unsupervised and thus not limited by training annotations, often the first limitation to the use of deep learning methods in the analysis of biomedical images. To our knowledge, this is the first use of cycle-GANs in a multi-task framework. Additionally, while generally used for synthetic images generation and data-augmentation, here we employ the cycle-GAN algorithm as an image-denoiser and cell-identifier. Indeed, when using cycle-GANs for data augmentation, we often want to learn the noise model of the image. Here, we make use of the cycle-consistency property of cycle-GANs and use the denoising model learned in parallel to the noise model to segment our images.

We compare the proposed method to five other approaches, a supervised method based on a U-net architecture, a pre-trained StarDist applied to Gabor-filtered images, two unsupervised

approaches based on a cycle-GAN with different inputs, and the FIAP.

We demonstrate that the presented DermoGAN architecture performs on par with expert manual identification of cells and outperforms the five other tested automated methods in accuracy and execution computational time. We also show that DermoGAN can be extended, with no retraining, to images acquired using other image acquisition techniques, and can be trained on datasets made entirely of synthetic images and deliver good results.

4.1 Methods

4.1.1 Identifying keratinocytes on RCM images with DermoGAN

The goal of the proposed DermoGAN model, shown in Fig. 24, is to estimate a mapping G_{A2B} from an RCM image domain (A) towards a binary domain (B). The mapping is learned using two connected complementary tasks, the first one learns RCM images noise and texture model (the likelihood of the image) from two sets of unpaired images: a set of RCM images and a set of (synthetic) binary images (obtained by simulating a prior model). The second task maps Gabor-filtered RCM images (domain C), *i.e.*, where membranes have been highlighted and the noise removed, into binary (synthetic) images, to learn the global geometrical structure of epidermal cell populations. The combination of the two tasks makes the overall model structure-aware, allowing us to denoise RCM images while highlighting membrane positions.

The proposed architecture is fully unsupervised, thus circumventing the obstacle of limited labelled data. Additionally, as it does not rely on training with a manually generated ground truth, as opposed to supervised approaches like U-net, its accuracy cannot be impacted by incorrectly labelled data, *i.e.*, missing cells in the ground truth or wrong detections.

Each task is a cycle-GAN network, made of two generators, denoted G_{A2B} and G_{B2A} in the first task and G_{C2B} and G_{B2C} in the second task, and two discriminators, denoted D_{B_1} and D_A in the first task and D_C and D_{B_2} in the second task, making a total of 8 networks in the model.

Generator and Discriminator architecture

The generator and discriminator networks form pairs (G_{A2B}/D_{B_1} and G_{B2A}/D_A , G_{C2B}/D_{B_2} and G_{B2C}/D_C). A generator takes a 256x256 image as input, down-samples it to extract high-level features and reduce spatial resolution, applies a succession of residual (ResNet) blocks to these features, and then up-samples them to increase the spatial resolution back up and generate the output, as described in Fig. S3. Each generator aims to create realistic target images taking a source image as input. The generators are constrained by an identity loss²⁰³ to ensure that the generator does not modify a target domain image, if used as an input, encouraging it to be an identity mapping, *i.e.*, $G_{A2B}(B) \approx B$. The two generators in the network should be cycle consistent to ensure that the data is preserved during the translation process and that the latter is reversible, *i.e.*, $G_{A2B}(G_{B2A}(B)) \approx B$ ²⁰⁴.

The weights in all generators were initiated using a Xavier (or Glorot) normal distribution²⁰⁵ such that the variation of the activations are the same across all layers to reduce the risk of the gradient exploding or vanishing and is a random number with a normal probability distribution in the range $\pm \sqrt{\frac{6}{n_i + n_o}}$, where $n_i = 862$ is the number of input images (both real RCM images and Gabor-filtered ones), and $n_o = 400$ is the number of output images (synthetic binary images). The weights of the generators were then updated by minimizing 3 loss functions (see Fig. S3).

In the case of G_{A2B} , these loss functions are: (1) An adversarial loss calculated with a mean squared error (MSE) between the generator and its associated discriminator, here D_{B_1} , such that for a pixel at coordinates $[i, j]$ of the generated image Gen_B , it is defined as,

$$MSE(D_{B_1}(Gen_B), 1) = \frac{1}{n_t} \sum_{i,j}^{n_t} (D_{B_1}(Gen_B)(i, j) - 1)^2 \quad (13)$$

where n_t is the size of the tensor outputted by the discriminator.

(2) An identity loss with a mean absolute error (MAE) between the input image I_B from domain B and the theoretical identity mapping $Id_{I_B} = G_{A2B}(I_B) \approx I_B$, defined for image I_B at pixel $[i, j]$,

$$MAE(Id_{I_B}, I_B) = \frac{1}{n_i} \sum_{i,j}^{n_i} |Id_{I_B}(i, j) - I_B(i, j)| \quad (14)$$

(3) A (Forwards or backwards) cycle consistency loss with a MAE between an input image I_B from domain B and the corresponding reconstructed image $Rec_{I_B} = G_{A2B}(G_{B2A}(I_B))$, defined at pixel of coordinates $[i, j]$ as,

$$MAE(Rec_{I_B}, I_B) = \frac{1}{n_i} \sum_{i,j}^{n_i} |Rec_{I_B}(i, j) - I_B(i, j)| \quad (15)$$

This loss function participates 10 times more to the update of the generator weights compared to the adversarial MSE loss.

The generators were trained with the ADAM optimizer with an initial learning rate of 0.002, and a decay rate of the gradient exponential moving average of 1.

The discriminators take an image as input and output the classification results (real vs. fake) in a tensor. Each discriminator aims to distinguish between real and generated target images, thus working against its matching adversary generator, which aims to create indiscriminable

generated target images. These two networks are connected through the adversarial loss in (13)²⁰⁴, and the discriminator loss function is,

$$\frac{1}{2}MSE(D_{B_1}(Gen_B), 1) + \frac{1}{2}MSE(D_{B_1}(I_B), 0) \quad (16)$$

Training each generator/discriminator pair simultaneously allows the cycle-GAN to learn the bidirectional image-to-image translation between two unpaired domains.

Multi-task approach

RCM images are noisy and heterogeneous due to tissue-induced scattering²⁰⁶, and are non-specific to organelles and macro-structures. This makes the identification of keratinocytes on RCM images a challenging task, whether done manually or automatically. In this case, cell identification requires two simultaneous tasks to capture the breath of information in confocal images: noise removal and membrane identification. Multi-task learning allows for concurrent execution of these two related tasks, improving overall performance by leveraging complementary information and sharing representations²⁰⁷. This reasoning mimics the human expert's approach to manual cell identification on RCM images, *i.e.*, focusing on bright tube-like membranes while ignoring the bright blob-like noise.

Noise removal was performed using a first cycle-GAN network, learning the translation between RCM images and binary images obtained by simulating a prior model, whereas membrane identification was performed by learning the mapping between binary images and Gabor-filtered RCM images, *i.e.*, where membranes were highlighted.

The multi-task model is optimized through soft-sharing of parameters²⁰⁸, since the two tasks do not share any hidden layer but are connected through their loss function, as shown in Fig. 24. Indeed, at each update of the loss functions, those associated with the generators creating the binary images, *i.e.*, G_{A2B} and G_{C2B} , are updated through their regular optimization, and then the

maximum value of the two trios of losses is set as the loss function for both generators, in order to synchronize training across the two tasks of noise removal and membrane identification.

4.1.2 Refining the results with star-convex polygons detection

The proposed method is a top-down approach to cell detection. DermoGAN roughly localizes individual cell locations, but post-processing is required, as shown in Fig. 25 and Fig. S4. Indeed, applying the obtained mapping G_{A2B} to a locally normalized split RCM image ($1000 \times 1000 \rightarrow 256 \times 256$) results in an incomplete binary image. To guarantee that the outside contour of tissue where the keratinocytes are detected is closed, we compute the alpha shape²⁰⁹ of the incomplete binary mask at a set level of refinement, such that the tissue comprises only one volume per external contour and is not broken down into smaller shapes, and that the alpha-shape contour matches the actual tissue area. Alpha-shapes are a generalization of convex hulls and are a mathematical approach formalizing the concept of a shape encompassing a set of points. Small holes in the membrane are then closed using a connected components analysis. We assume that all cells are star-convex shapes. However non-star-convex polygons can result from the false merging of two or more cells due to the lack of contrast on the membranes. To split these shapes, we use the pre-trained CNN StarDist²¹⁰ to detect star-convex polygons within the contours detected by the DermoGAN model, consequently refining our results, countering any missed cells and reducing the number of false negatives, as shown in Fig. 24.

4.2 Experiments and Results

4.2.1 Dataset

Images used in DermoGAN, U-net, and both approaches using a cycle-GAN, were of size 256×256 pixels and obtained by splitting the full RCM image (described in Chapter 3) into 9 non-overlapping square patches of 256×256 pixels. The FIAP network used full RCM images.

The results will be compared to the ground truth which was generated by Expert 1 by manually pointing out cell centers on 9 RCM images of 7 subjects, aged 5 months to 35 years.

Generated simulations. Binary images of size 256x256 pixels were created by the previously described hard-core process.

Synthetic RCM images were generated by adding different levels of noise and brightness heterogeneity to the binary images, as shown in Fig. S2.

4.2.2 DermoGAN implementation details

The model was trained for 5172 epochs on 46.9 CPU cores and 85.4 GiB of RAM. Training took approximately 4 days. All deep learning models were implemented using PyTorch.

4.2.3 Comparison to other automated methods

The proposed method is compared to 5 other approaches: (1) a deep learning approach based on the U-net architecture, (2) a StarDist algorithm applied to Gabor-filtered RCM images, (3) a cycle-GAN trained to translate RCM images into binary images, (4) a second cycle-GAN trained to turn Gabor-filtered RCM images into binary images, and (5) the FIAP^{151,153,211}.

The proposed combination of cycle-GAN models into a multi-task approach improves results by mimicking manual expertise, disregarding noise to focus on membrane location and tissue structure.

U-net. The U-net architecture which outperformed the other machine learning methods presented in Chapter 3 was selected to be compared to DermoGAN. It was pre-trained on the 2012 ImageNet Large Scale Visual Recognition Challenge dataset²¹² with an efficientnetb3 backbone¹⁹⁴, and was further trained on 43 real RCM images (4 participants, aged 20 – 35 years) and 203 synthetic RCM images and tested on 13 real RCM images and 68 synthetic RCM images. The network used a combination of two loss functions: Dice loss¹⁹⁵ and Focal loss¹⁹⁶ to account for class imbalance between cell membranes and background. The model is trained

with the ADAM optimizer with an initial learning rate of 0.0001, a batch size of 64, and a sigmoid activation function. The model was trained for 500 epochs on 46.9 CPU cores and 85.4 GiB of RAM. Training took approximately three days.

StarDist applied to Gabor-filtered images. A Gabor-filter was applied to ROI-masked RCM images to highlight membrane positions. The result was normalized with a histogram equalization and binarized with a Gaussian adaptive thresholding. A pre-trained StarDist was then applied to the binary masked-Gabor-filtered RCM image.

Cycle-GAN based models. Two cycle-GAN models were trained on 647 RCM images and 276 binary images, each one representing a task in the DermoGAN architecture, to evaluate each model independently, and later emphasize the importance of combining the two tasks into one architecture. The first one aimed to translate RCM images into binary images, whereas the second sought to convert Gabor-filtered RCM images into binary images. Both tested cycle-GANs models were refined using star-convex shape detection as performed in the DermoGAN architecture. Training was performed for 12068 epochs on 46.9 CPU cores and 85.4 GiB of RAM and took two days.

FIAP. The previously presented 3-step pipeline for keratinocytes detection^{151,153,211} based on membrane detection using image filters was applied to full RCM images of size 1000x1000 pixels and compared to the other methods.

4.2.4 Keratinocytes identification results

The proposed DermoGAN architecture was evaluated using 9 full RCM images, each divided into 9 patches. Accuracy (precision and recall summarized into the F1-score) was calculated using d-accuracy¹⁶⁹ against a manually obtained ground truth and compared to results obtained with the six described methods, as shown in Fig. 26 and Tables 10 and S2.

The poor performance of the pre-trained U-net model augmented with real and synthetic RCM images with Focal and Dice loss functions is in part due to the limited training set. Being a supervised approach, it may also suffer from missing cells in the ground truth used for training, and from membranes in the ground truth images created by Voronoi tessellation initiated from manually determined cell centers, not matching the actual membrane position in RCM images. The pre-trained StarDist applied to Gabor-filtered images also performs poorly. Indeed, although the Gabor filter highlights most membranes, it may also highlight noise, due for example to organelles, leading to false positives and low precision. Although the StarDist post-processing greatly improves results by segmenting (correctly or not) the detected contours into star-convex shapes, it does not manage to correct for all missing cells, leading to false negatives and consequently low recall, and overall low F1-score.

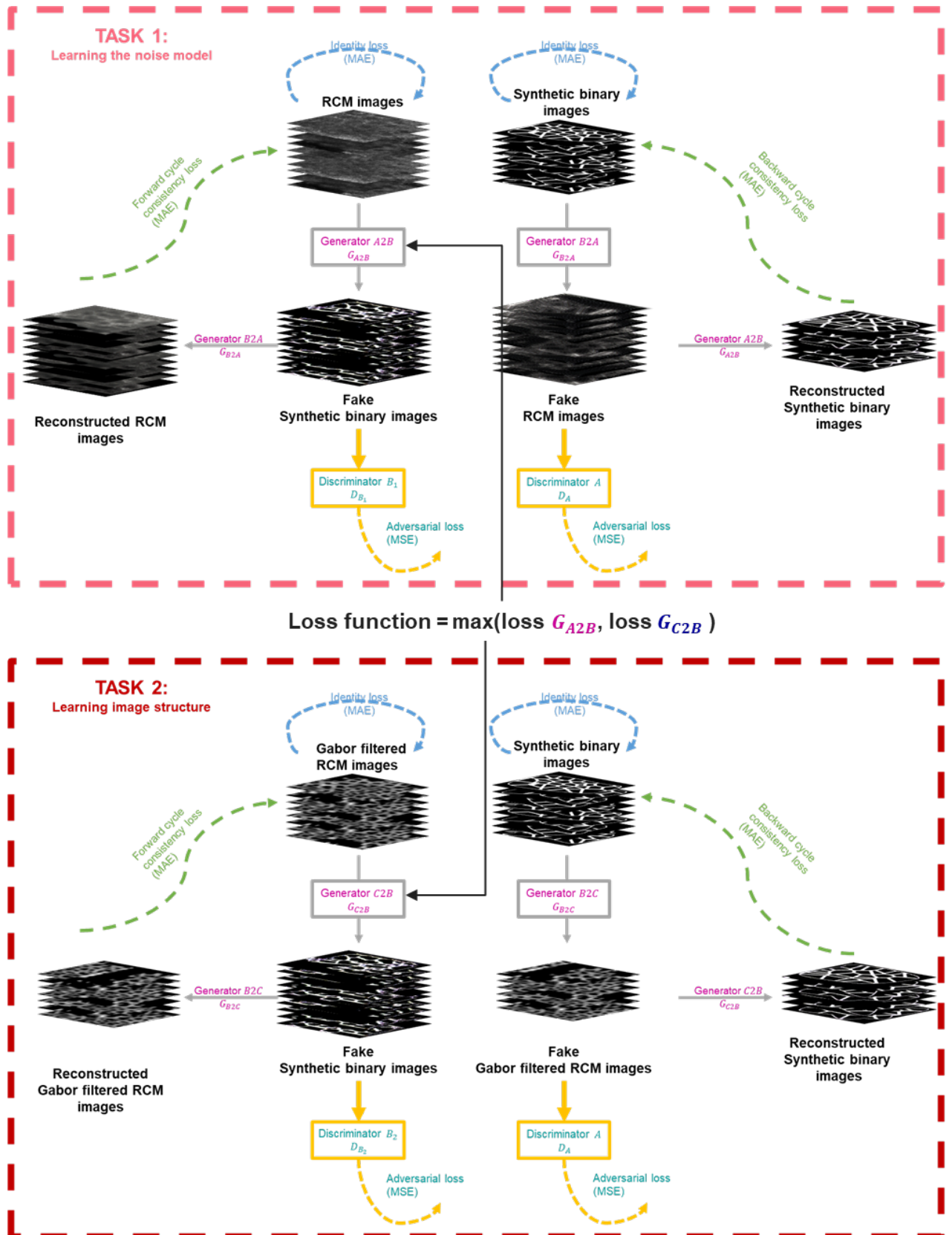


Fig. 24 DermoGAN architecture. The first task maps RCM images to the unpaired synthetic binary images. Whereas the second task learns the structure RCM images of the epidermis by translating Gabor filtered RCM images into binary images. RCM, Reflectance confocal microscopy.

Both cycle-GAN-based approaches have low F1-scores as they fail to detect complete membranes, as shown in Fig. 25. Indeed, the cycle-GAN model trained on RCM images and binary images, struggles to distinguish between noise and microstructures making up the membranes. On the other hand, the cycle-GAN trained on Gabor-filtered images with binary images is corrupted by the spatial correlation of noise and fails to detect any structure present in the image, as seen in Fig. 25, which also hints at the reason behind the better performance of DermoGAN. Indeed, it seems that adding up the two independent cycle-GAN outputs, would close most holes in the detected membranes by focusing on membrane detection and omitting any noise visible in them.

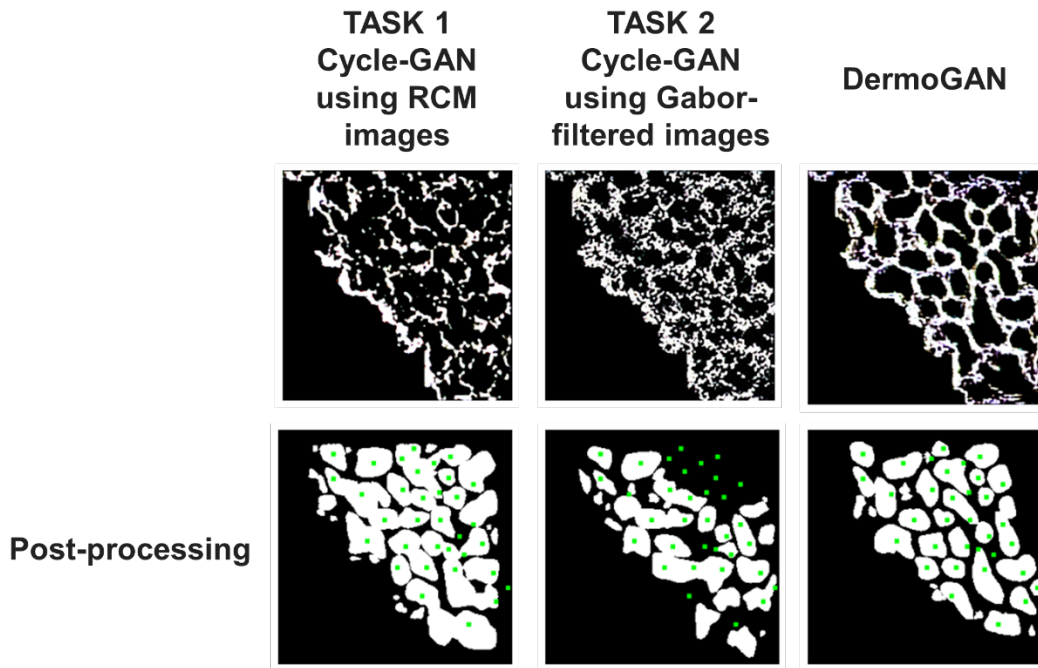


Fig. 45 Comparison of the two cycle-GAN based approaches and the proposed DermoGAN. Manually obtained ground truth in the form of cell centers in green. DermoGAN outperforms both methods.

Both DermoGAN and the FIAP outperform the other models, as shown in Tables 10 and S2, and show a great trade-off between precision and recall. DermoGAN has a higher F1-score than FIAP for 6 out of 9 images. The first seems to favor recall and is less likely to miss existing

cells and produce false negatives, whereas the second seems to favor precision and is less likely to invent cells and create false positive detections.

The DermoGAN architecture does not require manual parametrization nor a different set of parameters per epidermal layer, contrarily to the FIAP. This argues in favor of the DermoGAN architecture since multiple epidermal layers are often present in one RCM image. Once trained, its execution time is faster. It is based on the discovery of potentially unknown patterns in the image, making it less explainable than the FIAP. The latter is built on membrane detection using tubeness filters, with all its parameters being determined using general prior knowledge on the morphological features of the studied tissue. It is well documented that keratinocyte area increases with age and differs from one body site to another, and thus general parameters determined on a specific dataset may not be appropriate for all images. This point favors the DermoGAN architecture as more adaptable to different datasets and potentially to different image acquisition techniques and/or observed tissue.

Table 10 Comparison of median F1-score (computed with d-accuracy¹⁶⁹) for all six tested approaches.

	U-net based architectur e	StarDist on Gabor- filtered images	Cycle-GAN using RCM images	Cycle-GAN using Gabor- filtered images	FIAP	DermoGA N
Median F1- score	47,9	38,9	41,6	30,4	65,1	69
Standard deviation	12,2	12,1	6,4	8,3	7,4	4,2

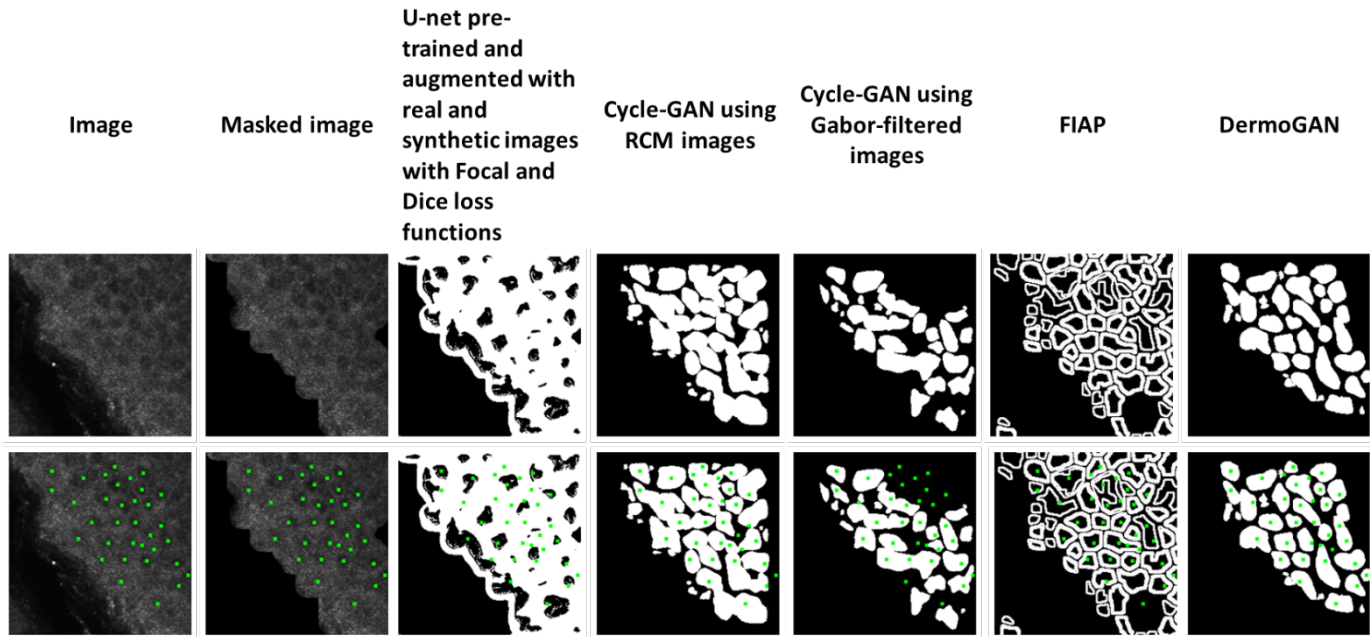


Fig. 26 RCM image analyzed with the 5 presented algorithms. DermoGAN outperforms 4 out of 5 other methods on all images, and outperforms FIAP on 6 out of 9 images. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

This study only included Caucasians with Fitzpatrick skin phototypes from I to III. Nevertheless, we tested DermoGAN on two RCM images of the granular layer of subjects with heavily pigmented skin and compare the results to a manual ground truth obtained by Expert 1 (see Table S3). The results show low precision and high recall, *i.e.*, DermoGAN tends to over-segments cells on RCM images of heavily pigmented skin, probably due to the detection of melanin caps. The FIAP outperforms DermoGAN in the detection of keratinocytes on RCM images of heavily pigmented skin.

4.2.5 Extension to other image acquisition techniques

Although the presented method was trained using RCM images for the detection of keratinocytes, we hypothesized that it can be extended without retraining to images generated by other instruments. Indeed, multi-task learning methods tend to perform well on domain adaptation and generalization and are therefore less data *dependent*. On the other hand, such

adaptability can lower pixel-level segmentation, and therefore is more suited when the accuracy is calculated at object-level and not at pixel-level.

To test this hypothesis, we applied the presented DermoGAN trained on RCM images on fluorescence microscopy images and compared the obtained results to 17 thresholding methods using ImageJ (detailed in the Appendix), as shown in Fig. 27. We observed that, although trained on different images of a different tissue, DermoGAN managed to identify membranes while omitting the noisy background, and outperformed traditional thresholding methods. It is important to note that the fluorescence microscopy image considered does show a similar tissue organization, *i.e.*, cohesive tissue with cells *sharing* membranes, to RCM images of the epidermis. However, when tested on cell culture images where cells were not always confluent, we noticed a loss in accuracy when using DermoGAN for cell identification. We therefore trained a second model using a different prior for simulating binary images and a different filter to enhance contours, that will be referred to for simplicity as DermoGAN2, on images where cells were not confluent.

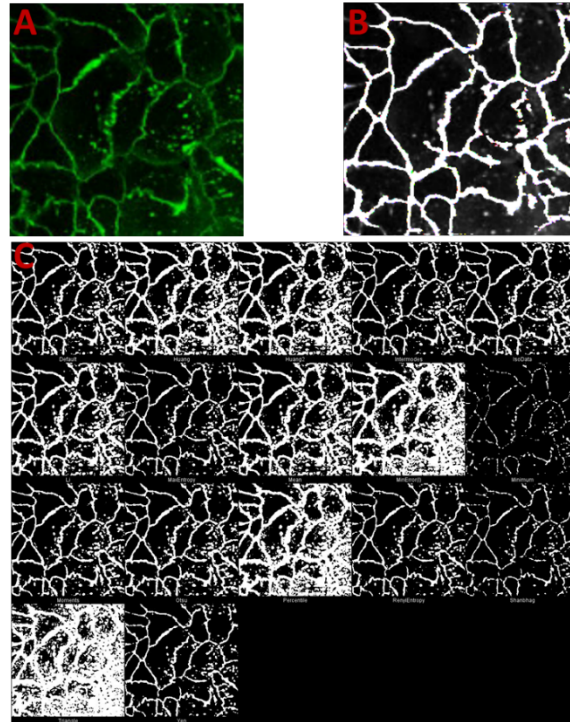


Fig. 27 DermoGAN can be extended using retraining to images acquired through different imaging techniques and outperforms traditional thresholding algorithms. **(A)** Input fluorescence microscopy images. **(B)** Output of DermoGAN applied to image A. **(C)** Application of 17 thresholding approaches to the same image.

4.2.6 Retraining the model with only synthetic images

DermoGAN2 was trained entirely on synthetic images. This served as a test of the generalization of the method when the available dataset is even more limited and serves to prove that the combination of the two tasks in the proposed model can capture general information and therefore can be extended to different images and tissues with similar organization, architecture, or texture, even when the images of interest were not included in the training set. The first task in DermoGAN2 maps synthetic non-confluent images created using the SIMCEP software for the simulation of fluorescence microscope images of cell populations²¹³ (Fig. 28A) to binary non-confluent images (Fig. 28C). The binary images were obtained by simulating a MPP embedding a constraint on overlap between objects defined by disks¹⁰². Meanwhile, the second task aims to learn the translation of Canny-filtered synthetic non-confluent images (Fig. 28B) towards the same binary non-confluent images.

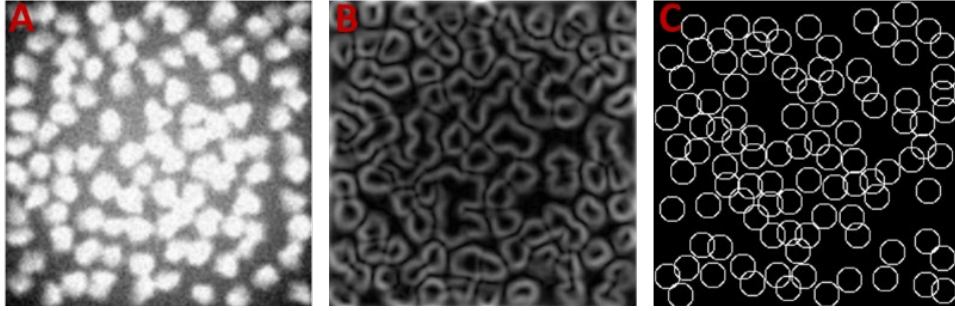


Fig. 28 DermoGAN2 was trained entirely on synthetic images. **(A)** Synthetic non-confluent images created using the SIMCEP software. **(B)** Canny-filtered non-confluent images created using the SIMCEP software. **(C)** Binary non-confluent images.

The resulting DermoGAN2 was then applied to images of cell cultures, and on mass spectroscopy images.

DermoGAN2 on cell culture images

We applied DermoGAN2 on an image of BV-2 microglial cells derived from C57/BL6 murine from the LIVECell dataset²¹⁴ as seen in Fig. 29. We obtained an accurate segmentation of the cells on the image. To avoid border effects in the image, a 10-pixels frame was applied to the image.

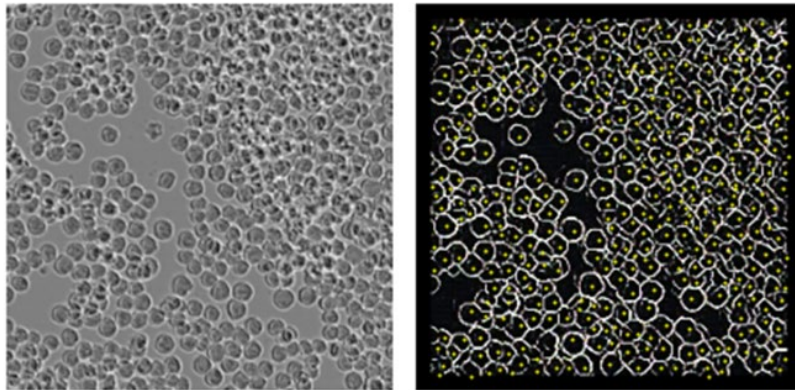


Fig. 29 DermoGAN2 applied to an image of confluent BV-2 cells (on the left), resulted in accurate detection of cells (on the right). Manually determined cell centers were plotted on DermoGAN2 output in yellow.

We also applied DermoGAN2 to an image of the SK-BR-3 human breast cancer cell line, where cells display morphological heterogeneity (Fig. 30A). Good cell detection was observed (Fig.

30B) on most cells when contrast is high enough. This proves that DermoGAN2 can be extended to different cell shapes and is not limited to the detection of the circular cells it was trained on, and that it is not restricted by the aspect of the cells. Indeed, the synthetic images used for training the model correspond to fluorescence images (bright cells in dark background resulting in high contrast) which is not the case of the tested cell culture images.

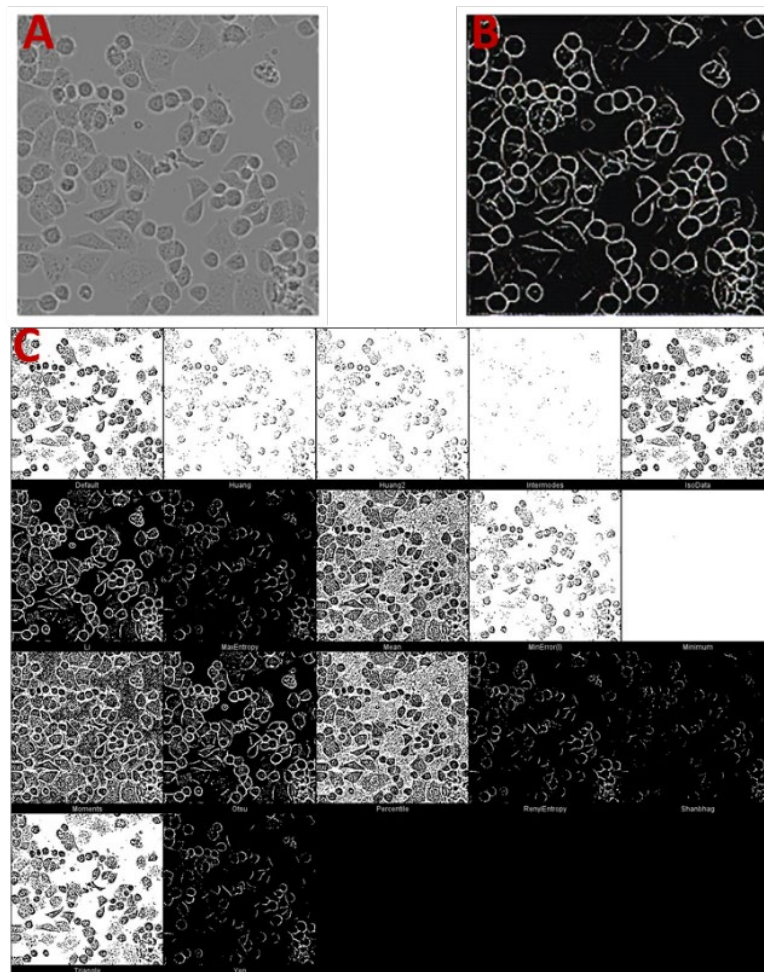


Fig. 30 DermoGAN2 applied to an image of SK-BR-3 cells (A) resulted in accurate detection (B) of cells compared to thresholding methods (C).

DermoGAN2 on mass spectroscopy images

Similar observations were made on mass spectroscopy images, where DermoGAN2 was able to detect hazy cell contours, with a tendency to merge neighboring cells into one detected

region, as shown in Fig. 31. This can be solved by post-processing using StarDist as done on RCM images.

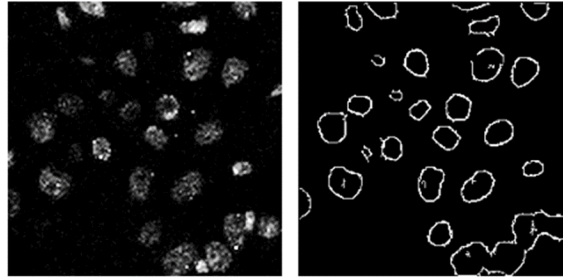


Fig. 31 DermoGAN2 applied to a mass spectroscopy image (on the left) resulted in accurate cell detection (on the right) with merging of adjacent cells.

The success of DermoGAN in segmenting cohesive tissues and that of DermoGAN2 in detecting non-confluent cells, highlights the importance of the binary denoised images domain (domain B in DermoGAN). This domain serves as a prior domain incorporating anterior knowledge in the model by describing the structure of the studied tissue. In DermoGAN, the prior is represented as a tissue island containing adjacent cells of similar size, while in DermoGAN2, this prior is represented by circular non-confluent cells. This prior domain summarizes our knowledge of the studied tissue and steers the training of the model towards the right solution. Therefore, to obtain the best results, the appearance of the tissue should guide the choice of the appropriate DermoGAN model based on the corresponding prior domain.

4.2.7 DermoGAN on histology images

A third DermoGAN model (DermoGAN3) was trained on histology images for the segmentation of nuclei that are a main feature in medical diagnosis from histology images.

The first task in DermoGAN3 maps the pre-processed histology images to synthetic images, representing nuclei, created using the SIMCEP software. While the second task, aims to learn the translation of Canny-filtered pre-processed histology images towards the same binary images.

Dataset

Histology images of healthy skin were acquired from the Clinical Proteomic Tumor Analysis Consortium Cutaneous Melanoma Collection of the Cancer Imaging Archive²¹⁵.

Each image of a total of 32 images contains one or multiple tissue slides of both dermis and epidermis and was stained with hematoxylin and eosin (H&E).

Creation of images of the epidermis. This work focuses on studying the architecture of the epidermis, we therefore identified the epidermis on each image. To do so, we applied a morphological geodesic active contour to the image and found the border between the epidermis and dermis. Indeed, the two layers are clearly distinguishable on histology slides, as the densely populated epidermis appears as a purple layer of connected tissue containing keratinocytes due to the binding of its structures to hematoxylin, while the more sparsely populated dermis appears as a pink irregular connective tissue due to the binding of eosin to the extracellular matrix and collagen, as visible in Fig. 32.

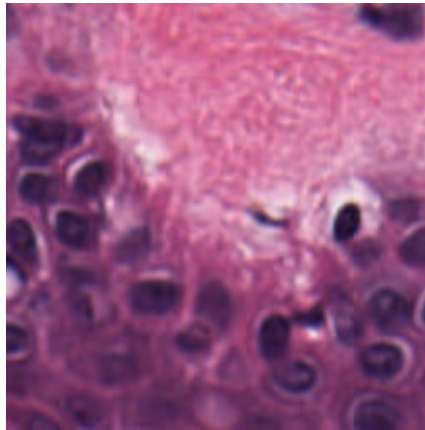


Fig. 32 Histology image of healthy skin. The epidermis appears as a purple connected layer, while the dermis appears as a sparsely populated pink tissue.

Color normalization. Color variation in histology images is common and is determined by the used scanner, equipment, and stain coloring manufacturers²¹⁶, hence the need for a color normalization method to reduce variances between images.

We used the Macenko stain normalizer²¹⁷ implemented in the torchstain package²¹⁸. It is an unsupervised normalization technique. First, the algorithm estimates the hematoxylin and eosin (H&E) stain vectors within the target image using a singular value decomposition (SVD) approach, specifically applied to the non-background pixels in the input image. Then, the algorithm applies a correction to compensate for intensity variations stemming from factors such as the original stain strength and the staining procedure. Lastly, the image is projected onto a reference image so that, following the stain normalization process, all the resulting images exhibit consistent color characteristics. Macenko's stain normalization is based on the premise that the color of each pixel can be expressed as a linear combination of two H&E stain vectors, which are initially unknown and must be estimated during the normalization process.

Results

We applied DermoGAN3 to a histology image of healthy epidermis which had previously been color normalized and observed that it successfully detects nuclei position., as shown in Fig. 33. We noticed however a tendency to rounding the detected shapes, which may be due to the prior used in forming the synthetic images.

Retraining the model was necessary here to introduce the information about the contrast between the nuclei and the cytoplasm, which was not present in the first two versions of the model. This approach was successful in detecting nuclei in histology images of healthy epidermis.

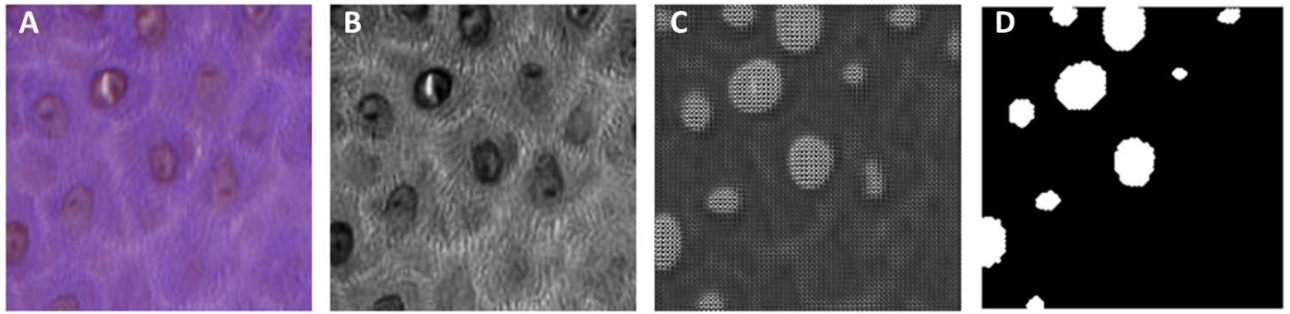


Fig. 33 Application of DermoGAN3 on a histology image of healthy skin. **(A)** Original histology image. **(B)** Color normalized image. **(C)** Output of DermoGAN3 using the color normalized image. **(D)** Binarized output of DermoGAN3.

4.3 Discussion & Conclusion

We have presented a novel multi-task cycle-GANs architecture for the identification of keratinocytes on RCM images which was compared to 4 other machine learning based methods and the FIAP. Supervised deep learning approaches obtained poor scores due to the lack of annotated data, even when using transfer learning. Unsupervised learning, such as cycle- GAN, failed to capture information at different scales simultaneously. Therefore, the FIAP approach outperformed these attempts. However, the proposed DermoGAN which combines two cycle-GANs to embed both local and global structural information, outperformed the classical FIAP, in terms of accuracy and execution time.

Cycle-GAN generators can encounter a vanishing gradient problem when the discriminator is too accurate in discriminating between input images and generated ones and therefore does not provide enough information for the optimization of the generator. This can be due to the images of the two domains being too different. To counter this issue, ResNet blocks could be replaced by Swin transformer blocks²¹⁹, which would also work better when using high-resolution images or images containing objects of different scales. Swin (Shifted Windows) transformer blocks use a shifted window approach, dividing images into non-overlapping patches and

processing them sequentially, shifting both horizontally and vertically, thus capturing information from different parts of the image. They use local and global self-attention computation, and so can capture both local and global (intra- and inter- patch) context efficiently. They also are convolution free, an advantage which would solve the checkboard effect sometimes observable on cycle-GANs generated images.

The current DermoGAN model is restricted to images of size 256x256, it would be useful to create a model applicable to images of all sizes, as is the case for StarDist.

We showed that the proposed fully unsupervised architecture can be used with or without retraining on other types of imaging and tissue types, bypassing the problem of required annotated data and potential label noise/missing labels in the ground truth. Moreover, it is not limited by the training set but rather determined by the prior data domain, *i.e.*, tissue organization and architecture in the training set. It would be of interest to generate new prior domains using MPPs to generate more specific priors, which would help extend the use of DermoGAN to images with multiple cell types or when the spatial dependence between different structures is important.

Chapter 5 Conclusion and Perspectives

In this thesis, we focused on the development of new approaches for the detection of keratinocytes in RCM images of the SG and SS of the human epidermis. Our motivation is the streamlining of this analysis for faster, more reproducible, and more unbiased detection of cells on these images. This could help spread the use of RCM in clinical practice and research. This work made use of computer vision techniques based on traditional image analysis methods, deep learning, and multi-task learning, and was extended to images acquired through different imaging techniques. In the following, we summarize the contributions made in this thesis and discuss future perspectives.

5.1 Summary of the contributions

FIAP. We proposed a three-step FIAP for the automated detection of keratinocytes on RCM images of the SG and SS. We identified the region of interest containing the epidermal cells, and then detected the individual cells in the segmented tissue area using tubeness filters to highlight membranes. We used prior biological knowledge on cell size to process the resulting detected contours, removing cells that are too small and reapplying the used filters locally on detected regions that are too big to be considered as a single cell.

The proposed FIAP was compared to machine learning-based approaches (Cell Cutter, and different U-net configurations and loss functions). These machine learning approaches showed less accuracy than the proposed method based on morphological features detection. This reflects the importance of a big (annotated) dataset when using machine learning algorithms, which is difficult to achieve in the case of RCM images, and biomedical images in general, reflecting the importance of, and need for either traditional computer vision methods, or unsupervised deep learning algorithms.

The proposed FIAP was useful in validating insights on infant skin maturation and adult skin ageing by automating a step that was previously done manually and required long tedious work. Indeed, it was applied to over 5000 RCM images of the SG and SS of clinical study participants, aged 0 to 80 years, and demonstrated the dynamic maturation and ageing of skin throughout one's life. This was, to our knowledge, the first large scale study of RCM images, and was useful in validating previous biological knowledge of the evolution of epidermal geometry and topology with age, and extending it to a broader age range, two body sites and two epidermal layers.

The FIAP is limited by its manual parametrization (one set of empirically determined parameters per epidermal layer) and consequently limited by the presence of multiple epidermal layers per image. To counter these limitations, we proposed a novel multitask cycle-GAN architecture, named DermoGAN.

DermoGAN. To counter the limitations of the FIAP, we proposed the dual-task cycle-GAN architecture DermoGAN. The first task optimized the mapping between real RCM images and binary images thus learning the noise and texture model of RCM images, whereas the second task optimized the translation between Gabor-filtered RCM images and binary images, learning the epidermal structure visible on RCM images. The combination of the two tasks allowed one task to constrict the solution space of the other thus improving overall results. We refined our cell identification by applying the pre-trained StarDist algorithm to detect star-convex shapes, thus closing any incomplete membranes and separating neighboring cells. We demonstrated that the proposed fully unsupervised method successfully identifies keratinocytes on RCM images of the epidermis, with an accuracy on par with human expert cell identification and generally outperforms the FIAP, and the tested machine learning methods.

We showed that DermoGAN is less *data dependent* than other methods as it can be generalized and adapted to images acquired using different imaging techniques (fluorescence microscopy, cell culture, mass spectroscopy and histology). This domain adaptability allowed us to train a second iteration of DermoGAN on solely synthetic images and a third on histology images, showing that the method can bypass issues of limited data set, and that domain adaptability and accuracy are dependent on the represented tissue organization and architecture in the data not on the nature of the data itself. Such adaptability can lead to a decrease in pixel-level segmentation accuracy, and therefore is more adapted when accuracy is calculated at object-level.

We tested DermoGAN and a cycle-GAN model for unpaired domain translation from RCM to histology and *vice versa*. Dermatologists are more often trained to read histology slides than RCM images and it would therefore be beneficial to have a method to leverage insights provided by RCM images in the form of biopsy slides. This explorative attempt had limited accuracy due to limited datasets (see Appendix) but is promising and warrants more exploration.

The proposed methods have their own advantages and limitations (see Table 11), and the preferred solution depends on the desired output. All the proposed methods were integrated as tools in a ready-to-use graphical user interface (GUI) to facilitate their use for RCM segmentation and interpretation.

Table 11 Comparison between the two main proposed methods for keratinocytes detection and manual cell identification.

	Advantages	Limitations
Manual cell identification	<ul style="list-style-type: none"> Fully explainable. 	<ul style="list-style-type: none"> Time consuming. Subjective.
Full Image Analysis Pipeline	<ul style="list-style-type: none"> Explainable: Based on knowledge of the morphological properties of the studied structures. Good accuracy against manual segmentation by expert graders. Satisfactory trade-off between recall and precision. 	<ul style="list-style-type: none"> Presence of multiple layers. Manual parametrization.
DermoGAN	<ul style="list-style-type: none"> Task-level explainability: comparative study of DermoGAN vs. each task independently shows the role played by each task. Faster than FIAP (excluding training time). Good accuracy against manual segmentation by expert graders. Satisfactory trade-off between recall and precision. Outperforms the FIAP in most cases. Domain adaptability and generalization to other images. 	<ul style="list-style-type: none"> Decision-level explainability: unclear how predictions are made within each task. Required post-processing due to lower pixel-level accuracy. Potentially due to generalization capabilities of multi-task networks.

5.2 Implementation in a ready-to-use GUI for cell segmentation

A GUI was developed including all the proposed methods: (1) FIAP for SG images, (2) FIAP for SS images, (3) DermoGAN trained on RCM images, (4) DermoGAN2 trained on synthetic images, (5) DermoGAN for RCM → histology translation, (6) DermoGAN for histology → RCM translation, (7) cycle-GAN for RCM → histology translation, and (8) cycle-GAN for histology → RCM translation. Roadblocks were set within the GUI to ensure that images of the

right size and format were used with the chosen algorithm (see Fig. 33). Table 12 lists the outputs of each method.

Table 12 Summary of the outputs of each method.

Method	Saved outputs
FIAP	<ul style="list-style-type: none"> • ROI mask • Image result of the application of Gabor and Sato filters and the corresponding skeleton overlayed to the original image • Cell contours and cell labels, on independent images and overlayed to original image • Table of cell areas • Table of cell centers coordinates
DermoGAN	<ul style="list-style-type: none"> • Image result generated by DermoGAN • Image result after StarDist application • Table of cell areas • Table of cell centers coordinates

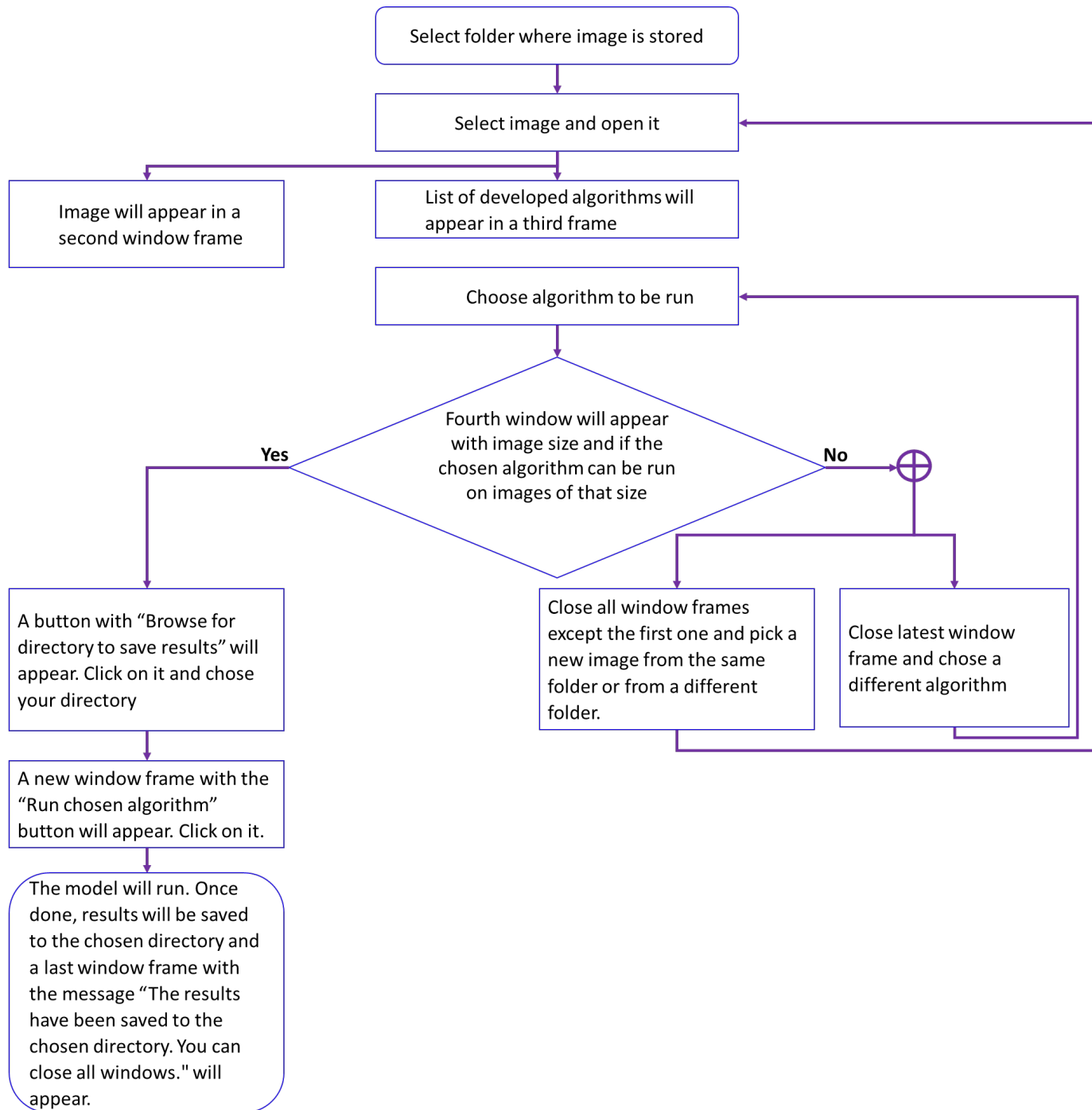


Fig. 34 Flowchart for the use of the developed graphical user interface including all developed methods.

5.3 Perspectives

The developed methods have only been evaluated on RCM images of healthy skin and showed accuracy on par with manual cell identification by human experts. It would be of interest to test

our methods on RCM images of diseased skin to evaluate the epidermal changes in structure, architecture, and topology of different skin conditions. This assessment may be a way to determine markers of disease progression based on quantitative cell characteristics which may be used to predict the onset of the disease, grade it, and evaluate its progression and reaction to treatment. It would also be interesting to validate the methods on image of heavily pigmented skin, as their application to these images has been so far limited.

A recent advancement in skin imaging is line-field confocal optical coherence tomography (LC-OCT) which has been commercially available since 2021. It combines the advantages of RCM (higher resolution than OCT) and OCT (higher penetration depth than RCM)²²⁰, and can be used in the diagnosis of cancer and other skin diseases^{221,222}. It would be of interest to test and maybe transpose the developed methods on images obtained by LC-OCT to evaluate the accuracy of keratinocyte detection on these images. Additionally, LC-OCT allows the visualization of cell nuclei (as dark areas in a brighter cell plasma background), which may be useful in the study of diseases with abnormal nuclei. Moreover, LC-OCT offers both horizontal and vertical views, making it a more adapted technique for RCM to histology unpaired translation, as the images from both domains could have the same orientation at more comparable depths (possible with OCT, but not with RCM), with cellular resolution (possible with RCM, but not with OCT). It would therefore be interesting to test the approaches developed in this thesis on LC-OCT images of the epidermis.

From a methodological point of view the proposed dual tasks cycle-GAN architecture open a new avenue to combining deep learning and Bayesian approaches. Indeed, by simulating a prior and a posterior we can create a database to train the model and thus avoid having to resort to a predefined database possibly with annotations, which is the main bottleneck of deep learning approaches today. Future investigations will be necessary to consider likelihoods associated

with different probes and priors proposed in the Bayesian literature for detecting objects such as the MPP framework.

Bibliography

1. T. Gambichler et al., "In vivo data of epidermal thickness evaluated by optical coherence tomography: Effects of age, gender, skin type, and anatomic site," *Journal of Dermatological Science* **44**(3), 145–152 (2006) [doi:10.1016/j.jdermsci.2006.09.008].
2. C. Longo et al., "Skin aging: in vivo microscopic assessment of epidermal and dermal changes by means of confocal microscopy," *J Am Acad Dermatol* **68**(3), e73-82 (2013) [doi:10.1016/j.jaad.2011.08.021].
3. J. Welzel et al., "Changes in function and morphology of normal human skin: evaluation using optical coherence tomography," *Br J Dermatol* **150**(2), 220–225 (2004) [doi:10.1111/j.1365-2133.2004.05810.x].
4. E. Sattler, R. Kästle, and J. Welzel, "Optical coherence tomography in dermatology," *J Biomed Opt* **18**(6), 061224 (2013) [doi:10.1117/1.JBO.18.6.061224].
5. T. Gambichler et al., "A comparative pilot study on ultraviolet-induced skin changes assessed by noninvasive imaging techniques in vivo," *Photochem Photobiol* **82**(4), 1103–1107 (2006) [doi:10.1562/2005-12-21-RA-757].
6. M. Rajadhyaksha et al., "In Vivo Confocal Scanning Laser Microscopy of Human Skin II: Advances in Instrumentation and Comparison With Histology," *The authors have declared conflict of interest., Journal of Investigative Dermatology* **113**(3), 293–303 (1999) [doi:10.1046/j.1523-1747.1999.00690.x].
7. M. Rajadhyaksha et al., "In vivo confocal scanning laser microscopy of human skin: melanin provides strong contrast," *J Invest Dermatol* **104**(6), 946–952 (1995) [doi:10.1111/1523-1747.ep12606215].
8. S. Gonzalez and Y. Gilaberte, "In vivo reflectance-mode confocal microscopy in clinical dermatology and cosmetology," *International journal of cosmetic science* **30**, 1–17 (2008) [doi:10.1111/j.1468-2494.2008.00406.x].
9. J. Bensaci et al., "Geometrical and topological analysis of in vivo confocal microscopy images reveals dynamic maturation of epidermal structures during the first years of life," *J Biomed Opt* **20**(9), 095004 (2015) [doi:10.1117/1.JBO.20.9.095004].
10. M. Huzaira et al., "Topographic variations in normal skin, as viewed by in vivo reflectance confocal microscopy," *J Invest Dermatol* **116**(6), 846–852 (2001) [doi:10.1046/j.0022-202x.2001.01337.x].
11. S. Guida et al., "Reflectance Confocal Microscopy of Aging Skin and Skin Cancer," *Dermatol Pract Concept* **11**(3), e2021068 (2021) [doi:10.5826/dpc.1103a68].
12. G. Stamatas and N. Kollias, "Infant skin maturation: Structural changes revealed by in vivo reflect," in *Reflectance Confocal Microscopy of Cutaneous Tumors*, 2nd ed. (2017).
13. M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning* **10**(4), 128–138 (1988) [doi:10.1002/sca.4950100403].
14. D. Gareau, Y. Patel, and M. Rajadhyaksha, "Basic principles of reflectance confocal microscopy," pp. 1–6 (2008) [doi:10.3109/9780203091562-2].
15. A. Levine and O. Markowitz, "Introduction to reflectance confocal microscopy and its use in clinical practice," *JAAD Case Rep* **4**(10), 1014–1023 (2018) [doi:10.1016/j.jdc.2018.09.019].
16. A. Raphael et al., "Reflectance Confocal Microscopy and Aging," pp. 1381–1397 (2017) [doi:10.1007/978-3-662-47398-6_161].
17. N. Kollias and G. N. Stamatas, "Optical non-invasive approaches to diagnosis of skin diseases," *J Invest Dermatol Symp Proc* **7**(1), 64–75 (2002) [doi:10.1046/j.1523-1747.2002.19635.x].
18. M. B. Murphrey, J. H. Miao, and P. M. Zito, "Histology, Stratum Corneum," in *StatPearls*, StatPearls Publishing, Treasure Island (FL) (2021).

19. A. Böhling et al., "Comparison of the stratum corneum thickness measured in vivo with confocal Raman spectroscopy and confocal reflectance microscopy," *1, Skin Res Technol* **20**(1), 50–57 (2014) [doi:10.1111/srt.12082].
20. M. Maarouf et al., "In Vivo Reflectance Confocal Microscopy: Emerging Role in Noninvasive Diagnosis and Monitoring of Eczematous Dermatoses," *8, Actas Dermosifiliogr* **110**(8), 626–636, Elsevier (2019) [doi:10.1016/j.adengl.2019.06.005].
21. G. Egawa, "Pathomechanism of 'skin-originated' allergic diseases," *4, Immunol Med* **41**(4), 170–176 (2018) [doi:10.1080/25785826.2018.1540257].
22. C. Longo, "Well-aging: Early Detection of Skin Aging Signs," *4, Dermatol Clin* **34**(4), 513–518 (2016) [doi:10.1016/j.det.2016.05.014].
23. J. L. S. Sánchez-Mateos et al., "Reflectance-Mode Confocal Microscopy in Dermatological Oncology," in *Lasers in Dermatology and Medicine*, K. Nouri, Ed., pp. 285–308, Springer, London (2012) [doi:10.1007/978-0-85729-281-0_22].
24. G. Raposo and M. S. Marks, "Melanosomes – dark organelles enlighten endosomal membrane transport," *10, Nat Rev Mol Cell Biol* **8**(10), 786–797 (2007) [doi:10.1038/nrm2258].
25. G. N. Stamatas et al., "Infant skin physiology and development during the first years of life: a review of recent findings based on in vivo studies," *1, Int J Cosmet Sci* **33**(1), 17–24 (2011) [doi:10.1111/j.1468-2494.2010.00611.x].
26. J. Nikolovski et al., "Barrier function and water-holding and transport properties of infant stratum corneum are different from adult and continue to develop through the first year of life," *7, J Invest Dermatol* **128**(7), 1728–1736 (2008) [doi:10.1038/sj.jid.5701239].
27. G. N. Stamatas et al., "Skin maturation from birth to 10 years of age: Structure, function, composition and microbiome," *Experimental Dermatology* **n/a**(n/a) [doi:10.1111/exd.14843].
28. G. N. Stamatas et al., "Infant skin microstructure assessed in vivo differs from adult skin in organization and at the cellular level," *2, Pediatr Dermatol* **27**(2), 125–131 (2010) [doi:10.1111/j.1525-1470.2009.00973.x].
29. E. Cinotti et al., "Structural skin changes in elderly people investigated by reflectance confocal microscopy," *11, J Eur Acad Dermatol Venereol* **34**(11), 2652–2658 (2020) [doi:10.1111/jdv.16466].
30. K. Kawasaki, K. Yamanishi, and H. Yamada, "Age-related morphometric changes of inner structures of the skin assessed by in vivo reflectance confocal microscopy," *3, Int J Dermatol* **54**(3), 295–301 (2015) [doi:10.1111/ijd.12220].
31. F. Alsamad and G. N. Stamatas, "Directional assessment of the skin barrier function in vivo," *Skin Research and Technology* **29**(5), e13346 (2023) [doi:10.1111/srt.13346].
32. S. G. Lagarrigue et al., "In vivo quantification of epidermis pigmentation and dermis papilla density with reflectance confocal microscopy: variations with age and skin phototype," *4, Exp Dermatol* **21**(4), 281–286 (2012) [doi:10.1111/j.1600-0625.2012.01451.x].
33. N. S. K. Haytoglu et al., "Assessment of skin photoaging with reflectance confocal microscopy," *3, Skin Res Technol* **20**(3), 363–372 (2014) [doi:10.1111/srt.12127].
34. S. González et al., "Characterization of psoriasis in vivo by reflectance confocal microscopy," *5–6, J Med* **30**(5–6), 337–356 (1999).
35. M. Ardigo et al., "Concordance between in vivo reflectance confocal microscopy and histology in the evaluation of plaque psoriasis," *J Eur Acad Dermatol Venereol* **23**(6), 660–667 (2009) [doi:10.1111/j.1468-3083.2009.03134.x].
36. M. Agozzino et al., "Monitoring treatment response in psoriasis: current perspectives on the clinical utility of reflectance confocal microscopy," *Psoriasis (Auckl)* **7**, 27–34 (2017) [doi:10.2147/PTT.S107514].
37. S. Astner et al., "Pilot study on the sensitivity and specificity of in vivo reflectance confocal microscopy in the diagnosis of allergic contact dermatitis," *Journal of the American Academy of Dermatology* **53**(6), 986–992 (2005) [doi:10.1016/j.jaad.2005.08.026].

38. S. Astner, S. González, and E. Gonzalez, "Noninvasive evaluation of allergic and irritant contact dermatitis by in vivo reflectance confocal microscopy," *4, Dermatitis* **17**(4), 182–191 (2006) [doi:10.2310/6620.2006.05052].
39. G. Pellacani et al., "Reflectance confocal microscopy made easy: The 4 must-know key features for the diagnosis of melanoma and nonmelanoma skin cancers," *2, J Am Acad Dermatol* **81**(2), 520–526 (2019) [doi:10.1016/j.jaad.2019.03.085].
40. N. Shahriari et al., "Reflectance Confocal Microscopy Criteria of Pigmented Squamous Cell Carcinoma In Situ," *3, The American Journal of Dermatopathology* **40**(3), 173–179 (2018) [doi:10.1097/DAD.0000000000000938].
41. I. Alarcon et al., "Impact of in vivo reflectance confocal microscopy on the number needed to treat melanoma in doubtful lesions," *British Journal of Dermatology* **170**(4), 802–808 (2014) [doi:10.1111/bjd.12678].
42. G. Pellacani et al., "Cost–benefit of reflectance confocal microscopy in the diagnostic performance of melanoma," *Journal of the European Academy of Dermatology and Venereology* **30**(3), 413–419 (2016) [doi:10.1111/jdv.13408].
43. A. Waddell, P. Star, and P. Guitera, "Advances in the use of reflectance confocal microscopy in melanoma," *Melanoma Management* **5**(1), MMT04, Future Medicine (2018) [doi:10.2217/mmt-2018-0001].
44. M. Rajadhyaksha et al., "Reflectance confocal microscopy of skin in vivo: From bench to bedside," *Lasers in Surgery and Medicine* **49**(1), 7–19 (2017) [doi:10.1002/lsm.22600].
45. R. Kong et al., "A comparative study of the effects of retinol and retinoic acid on histological, molecular, and clinical properties of human skin," *1, J Cosmet Dermatol* **15**(1), 49–57 (2016) [doi:10.1111/jocd.12193].
46. M. Manfredini et al., "Does skin hydration influence keratinocyte biology? In vivo evaluation of microscopic skin changes induced by moisturizers by means of reflectance confocal microscopy," *3, Skin Res Technol* **19**(3), 299–307 (2013) [doi:10.1111/srt.12042].
47. I. S. Bağcı et al., "Effects of Short-Term Moisturizer Application in Different Ethnic Skin Types: Noninvasive Assessment with Optical Coherence Tomography and Reflectance Confocal Microscopy," *3, SPP* **31**(3), 125–133, Karger Publishers (2018) [doi:10.1159/000486626].
48. N. de Carvalho et al., "15321 Evaluation of a ceramide-containing lotion on skin hydration and cellular morphology assessed by reflectance confocal microscopy," *6, Journal of the American Academy of Dermatology* **83**(6), AB147, Elsevier (2020) [doi:10.1016/j.jaad.2020.06.676].
49. C. Ruini et al., "In vivo examination of healthy human skin after short-time treatment with moisturizers using confocal Raman spectroscopy and optical coherence tomography: Preliminary observations," *1, Skin Research and Technology* **28**(1), 119–132 (2022) [doi:10.1111/srt.13101].
50. A. Gomes-Neto et al., "Efficacy of a Daily Protective Moisturizer with High UVB and UVA Photoprotection in Decreasing Ultraviolet Damage: Evaluation by Reflectance Confocal Microscopy," *10, Acta Derm Venerol* **97**(10), 1196–1201 (2017) [doi:10.2340/00015555-2736].
51. A. Bozkurt et al., "Skin strata delineation in reflectance confocal microscopy images using recurrent convolutional networks with attention," *1, Sci Rep* **11**(1), 12576 (2021) [doi:10.1038/s41598-021-90328-x].
52. A. Bozkurt et al., "Delineation of Skin Strata in Reflectance Confocal Microscopy Images using Recurrent Convolutional Networks with Toeplitz Attention," *arXiv:1712.00192 [cs]* (2017).
53. P. Kaur et al., "Hybrid deep learning for Reflectance Confocal Microscopy skin images," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 1466–1471 (2016) [doi:10.1109/ICPR.2016.7899844].
54. S. Hames et al., "Automated Segmentation of Skin Strata in Reflectance Confocal Microscopy Depth Stacks," *PLoS ONE* **11** (2016) [doi:10.1371/journal.pone.0153208].
55. E. Somoza et al., "Automatic Localization of Skin Layers in Reflectance Confocal Microscopy," *22 October 2014*, 141–150 [doi:10.1007/978-3-319-11755-3_16].

56. S. Kurugol et al., "Validation Study of Automated Dermal/Epidermal Junction Localization Algorithm in Reflectance Confocal Microscopy Images of Skin," *Proc SPIE Int Soc Opt Eng* **8207** (2012) [doi:10.1117/12.909227].
57. C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," *arXiv:1512.00567 [cs]* (2015).
58. S. Koller et al., "In vivo reflectance confocal microscopy: automated diagnostic image analysis of melanocytic skin tumours," *J Eur Acad Dermatol Venereol* **25**(5), 554–558 (2011) [doi:10.1111/j.1468-3083.2010.03834.x].
59. K. Kose et al., "A machine learning method for identifying morphological patterns in reflectance confocal microscopy mosaics of melanocytic skin lesions in-vivo," 968908 (2016) [doi:10.1117/12.2212978].
60. A. Bozkurt et al., "A Multiresolution Convolutional Neural Network with Partial Label Training for Annotating Reflectance Confocal Microscopy Images of Skin," *arXiv:1802.02213 [cs]* (2018).
61. A. Halimi et al., "An unsupervised Bayesian approach for the joint reconstruction and classification of cutaneous reflectance confocal microscopy images," 1 August 2017, 241–245 [doi:10.23919/EUSIPCO.2017.8081205].
62. S. Zorgui et al., "A Convolutional Neural Network for Lentigo Diagnosis," *The Impact of Digital Technologies on Public Health in Developed and Developing Countries* **12157**, 89–99 (2020) [doi:10.1007/978-3-030-51517-1_8].
63. D. Gareau et al., "Automated detection of malignant features in confocal microscopy on superficial spreading melanoma versus nevi," *J Biomed Opt* **15**(6), 061713 (2010) [doi:10.1117/1.3524301].
64. D. Gareau, "Automated identification of epidermal keratinocytes in reflectance confocal microscopy," 3, *J Biomed Opt* **16**(3), 030502 (2011) [doi:10.1117/1.3552639].
65. F. Lux and P. Matula, *Cell Segmentation by Combining Marker-Controlled Watershed and Deep Learning* (2020).
66. H. S. Wu, J. Barba, and J. Gil, "Iterative thresholding for segmentation of cells from noisy images," *J Microsc* **197**(Pt 3), 296–304 (2000) [doi:10.1046/j.1365-2818.2000.00653.x].
67. J. Qi, H. Yang, and Z. Kong, "A review of traditional image segmentation methods," 124511P (2022) [doi:10.1117/12.2656653].
68. N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," 1, *IEEE Transactions on Systems, Man, and Cybernetics* **9**(1), 62–66 (1979) [doi:10.1109/TSMC.1979.4310076].
69. S. P. Shen et al., "Automatic Cell Segmentation by Adaptive Thresholding (ACSAT) for Large-Scale Calcium Imaging Datasets," *eNeuro* **5**(5), ENEURO.0056-18.2018 (2018) [doi:10.1523/ENEURO.0056-18.2018].
70. A. Lagree et al., "A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks," 1, *Sci Rep* **11**(1), 8025, Nature Publishing Group (2021) [doi:10.1038/s41598-021-87496-1].
71. R. Harrabi and E. Ben Braiek, "Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images," *EURASIP Journal on Image and Video Processing* **2012**(1), 11 (2012) [doi:10.1186/1687-5281-2012-11].
72. S. Mayala and J. B. Haugsøen, "Threshold estimation based on local minima for nucleus and cytoplasm segmentation," *BMC Medical Imaging* **22**(1), 77 (2022) [doi:10.1186/s12880-022-00801-w].
73. T. Vicar et al., "Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison," *BMC Bioinformatics* **20**(1), 360 (2019) [doi:10.1186/s12859-019-2880-8].
74. W. Zhang and H. Li, "Automated segmentation of overlapped nuclei using concave point detection and segment grouping," *Pattern Recognition* **71**, 349–360 (2017) [doi:10.1016/j.patcog.2017.06.021].

75. M. Liao et al., "Automatic segmentation for cell images based on bottleneck detection and ellipse fitting," *Neurocomputing* **173**, 615–622 (2016) [doi:10.1016/j.neucom.2015.08.006].
76. I. Sobel and G. Feldman, "A 3×3 isotropic gradient operator for image processing," *Pattern Classification and Scene Analysis*, 271–272 (1973).
77. H. Scharr, "Optimal operators in digital image processing [Elektronische Ressource] /" (2014).
78. J. Canny, "A Computational Approach To Edge Detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8*, 679–698 (1986) [doi:10.1109/TPAMI.1986.4767851].
79. D. Marr and E. Hildreth, "Theory of edge detection," *Proc R Soc Lond B Biol Sci* **207**(1167), 187–217 (1980) [doi:10.1098/rspb.1980.0020].
80. Q. Wu and K. R. Castleman, "Chapter Seven - Image Segmentation," in *Microscope Image Processing (Second Edition)*, F. A. Merchant and K. R. Castleman, Eds., pp. 119–152, Academic Press (2023) [doi:10.1016/B978-0-12-821049-9.00003-4].
81. C. Garbay, J. M. Chassery, and G. Brugal, "An iterative region-growing process for cell image segmentation based on local color similarity and global shape criteria," *Anal Quant Cytol Histol* **8**(1), 25–34 (1986).
82. J. Ning et al., "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognition* **43**(2), 445–456 (2010) [doi:10.1016/j.patcog.2009.03.004].
83. D. Anoraganingrum, S. Kröner, and B. Gottfried, "Cell Segmentation with Adaptive Region Growing," 1999.
84. H. Su et al., "Automatic Myonuclear Detection in Isolated Single Muscle Fibers Using Robust Ellipse Fitting and Sparse Representation," *IEEE/ACM Trans Comput Biol Bioinform* **11**(4), 714–726 (2014) [doi:10.1109/TCBB.2013.151].
85. C. Panagiotakis and A. A. Argyros, "Cell Segmentation Via Region-Based Ellipse Fitting," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2426–2430 (2018) [doi:10.1109/ICIP.2018.8451852].
86. X. Bai, C. Sun, and F. Zhou, "Splitting touching cells based on concave points and ellipse fitting," *Pattern Recognition* **42**(11), 2434–2446 (2009) [doi:10.1016/j.patcog.2009.04.003].
87. J. Wu et al., *A novel color image segmentation method and its application to white blood cell image analysis* (2006) [doi:10.1109/ICOSP.2006.345700].
88. S. Tara et al., "Various Image Segmentation Methods Based On Partial Differential Equation-A Survey," 2248–9584 (2014).
89. J. Sun, "Application of Image Segmentation Algorithm Based on Partial Differential Equation in Legal Case Text Classification," *Advances in Mathematical Physics* **2021**, e4062200, Hindawi (2021) [doi:10.1155/2021/4062200].
90. S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics* **79**(1), 12–49 (1988) [doi:10.1016/0021-9991(88)90002-2].
91. T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing* **10**(2), 266–277 (2001) [doi:10.1109/83.902291].
92. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int J Comput Vision* **1**(4), 321–331 (1988) [doi:10.1007/BF00133570].
93. T. Atta-Fosu et al., "3D Clumped Cell Segmentation Using Curvature Based Seeded Watershed," *J Imaging* **2**(4), 31 (2016) [doi:10.3390/jimaging2040031].
94. T. Kaseva et al., "Marker-controlled watershed with deep edge emphasis and optimized H-minima transform for automatic segmentation of densely cultivated 3D cell nuclei," *BMC Bioinformatics* **23**(1), 289 (2022) [doi:10.1186/s12859-022-04827-3].
95. C. F. Koyuncu et al., "Iterative h-minima-based marker-controlled watershed for cell nucleus segmentation," *Cytometry A* **89**(4), 338–349 (2016) [doi:10.1002/cyto.a.22824].
96. B. Preim and C. Botha, Eds., "Visual Computing for Medicine," in *Visual Computing for Medicine (Second Edition)*, p. i, Morgan Kaufmann, Boston (2014) [doi:10.1016/B978-0-12-415873-3.00023-7].

97. Z. Wang, "Cell Segmentation for Image Cytometry: Advances, Insufficiencies, and Challenges," *Cytometry Part A* **95**(7), 708–711 (2019) [doi:10.1002/cyto.a.23686].
98. H. K. Hahn et al., "A Minimally-Interactive Watershed Algorithm Designed for Efficient CTA Bone Removal," in *Computer Vision Approaches to Medical Image Analysis*, R. R. Beichel and M. Sonka, Eds., pp. 178–189, Springer, Berlin, Heidelberg (2006) [doi:10.1007/11889762_16].
99. J.-M. Kuhnigk et al., "Lung lobe segmentation by anatomy-guided 3D watershed transform," *Proceedings of SPIE - The International Society for Optical Engineering* **5032** (2003) [doi:10.1117/12.480321].
100. K. Z. Mao, P. Zhao, and P.-H. Tan, "Supervised learning-based cell image segmentation for p53 immunohistochemistry," *IEEE Trans Biomed Eng* **53**(6), 1153–1163 (2006) [doi:10.1109/TBME.2006.873538].
101. M. Ghorbani et al., "Functional marked point processes: a natural structure to unify spatio-temporal frameworks and to analyse dependent functional data," *TEST* **30**(3), 529–568 (2021) [doi:10.1007/s11749-020-00730-2].
102. X. Descombes, "Multiple objects detection in biological images using a marked point process framework," *Methods* **115**, 2–8 (2017) [doi:10.1016/j.ymeth.2016.09.009].
103. M. Kulikova et al., "Nuclei Extraction from Histopathological Images using a Marked Point Process Approach," *Proc SPIE* **8314**, 75 (2012) [doi:10.1117/12.911757].
104. N. Gadgil et al., *Nuclei segmentation of fluorescence microscopy images based on midpoint analysis and marked point process*, p. 40 (2016) [doi:10.1109/SSIAI.2016.7459169].
105. Y. Li et al., "Point process models for localization and interdependence of punctate cellular structures," *Cytometry A* **89**(7), 633–643 (2016) [doi:10.1002/cyto.a.22873].
106. M. Ortner, X. Descombe, and J. Zerubia, "A marked point process of rectangles and segments for automatic analysis of digital elevation models," *IEEE Trans Pattern Anal Mach Intell* **30**(1), 105–119 (2008) [doi:10.1109/TPAMI.2007.1159].
107. A. Saxena et al., "A review of clustering techniques and developments," *Neurocomputing* **267**, 664–681 (2017) [doi:10.1016/j.neucom.2017.06.053].
108. M. Gamarra, "MC-Kmeans: an Approach to Cell Image Segmentation Using Clustering Algorithms," *International Journal of Artificial Intelligence* **19**, 14 (2021).
109. O. Sarrafzadeh and A. M. Dehnavi, "Nucleus and cytoplasm segmentation in microscopic images using K-means clustering and region growing," *Adv Biomed Res* **4**, 174 (2015) [doi:10.4103/2277-9175.163998].
110. C. Zhang et al., "White Blood Cell Segmentation by Color-Space-Based K-Means Clustering," *Sensors* **14**(9), 16128–16147, Multidisciplinary Digital Publishing Institute (2014) [doi:10.3390/s140916128].
111. O. Sarrafzadeh et al., "A simple and accurate method for white blood cells segmentation using K-means algorithm," in *2015 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 1–6 (2015) [doi:10.1109/SiPS.2015.7344978].
112. M. Yan et al., "K-means cluster algorithm based on color image enhancement for cell segmentation," in *2012 5th International Conference on BioMedical Engineering and Informatics*, pp. 295–299 (2012) [doi:10.1109/BMEI.2012.6513157].
113. W. Gonçalves and O. Bruno, "Automatic System for Counting Cells with Elliptical Shape," *Learn. NonLinear Models* **9** (2012) [doi:10.21528/LNLM-vol9-no1-art1].
114. S. S. Savkare and S. P. Narote, "Blood cell segmentation from microscopic blood images," in *2015 International Conference on Information Processing (ICIP)*, pp. 502–505 (2015) [doi:10.1109/INFOP.2015.7489435].
115. M. Abdolhoseini et al., "Segmentation of Heavily Clustered Nuclei from Histopathological Images," *1, Sci Rep* **9**(1), 4551, Nature Publishing Group (2019) [doi:10.1038/s41598-019-38813-2].
116. Y. LeCun et al., "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation* **1**(4), 541–551 (1989) [doi:10.1162/neco.1989.1.4.541].

117. G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis* **42**, 60–88 (2017) [doi:10.1016/j.media.2017.07.005].
118. J. P. Viguera-Guillén et al., "Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation," *BMC Biomedical Engineering* **1**(1), 4 (2019) [doi:10.1186/s42490-019-0003-2].
119. D. Cirean et al., "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images," *Proceedings of Neural Information Processing Systems* **25** (2012).
120. E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *4, IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(4), 640–651 (2017) [doi:10.1109/TPAMI.2016.2572683].
121. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]* (2015).
122. N. Siddique et al., "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications," *IEEE Access* **9**, 82031–82057 (2021) [doi:10.1109/ACCESS.2021.3086020].
123. O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas" (2018).
124. Z. Zhang et al., "DENSE-INception U-net for medical image segmentation," *Computer Methods and Programs in Biomedicine* **192**, 105395 (2020) [doi:10.1016/j.cmpb.2020.105395].
125. K. He et al., "Deep Residual Learning for Image Recognition" (2015) [doi:10.48550/arXiv.1512.03385].
126. M. Z. Alom et al., "Recurrent residual U-Net for medical image segmentation," *J Med Imaging (Bellingham)* **6**(1), 014006 (2019) [doi:10.1117/1.JMI.6.1.014006].
127. I. J. Goodfellow et al., "Generative Adversarial Networks," *arXiv:1406.2661, arXiv* (2014) [doi:10.48550/arXiv.1406.2661].
128. S. J. Gershman, "The Generative Adversarial Brain," *Front Artif Intell* **2**, 18 (2019) [doi:10.3389/frai.2019.00018].
129. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in *Computer Vision – ECCV 2016*, B. Leibe et al., Eds., pp. 694–711, Springer International Publishing, Cham (2016) [doi:10.1007/978-3-319-46475-6_43].
130. X. Mao et al., "Least Squares Generative Adversarial Networks," *arXiv:1611.04076, arXiv* (2017) [doi:10.48550/arXiv.1611.04076].
131. J. Kweon et al., "A Novel Method Based on GAN Using a Segmentation Module for Oligodendroglioma Pathological Image Generation," *Sensors (Basel)* **22**(10), 3960 (2022) [doi:10.3390/s22103960].
132. M. Häring et al., "Automated Segmentation of Epithelial Tissue Using Cycle-Consistent Generative Adversarial Networks," p. 311373, *bioRxiv* (2018) [doi:10.1101/311373].
133. D. Bähr et al., "Cellcyclegan: Spatiotemporal Microscopy Image Synthesis Of Cell Populations Using Statistical Shape Models And Conditional Gans," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 15–19 (2021) [doi:10.1109/ISBI48211.2021.9433896].
134. J. Jiang et al., "Tumor-Aware, Adversarial Domain Adaptation from CT to MRI for Lung Cancer Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi et al., Eds., pp. 777–785, Springer International Publishing, Cham (2018) [doi:10.1007/978-3-030-00934-2_86].
135. Y. Zhang and Q. Yang, "An overview of multi-task learning," *National Science Review* **5**(1), 30–43 (2018) [doi:10.1093/nsr/nwx105].
136. R. Caruana, "Multitask Learning," *Machine Learning* **28**(1), 41–75 (1997) [doi:10.1023/A:1007379606734].
137. Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," *IEEE Transactions on Knowledge and Data Engineering* **34**(12), 5586–5609 (2022) [doi:10.1109/TKDE.2021.3070203].

138. Y. Zhao et al., "Multi-task deep learning for medical image computing and analysis: A review," *Computers in Biology and Medicine* **153**, 106496 (2023) [doi:10.1016/j.combiomed.2022.106496].
139. S. Mukherjee et al., "Domain Adapted Multitask Learning for Segmenting Amoeboid Cells in Microscopy," *IEEE Transactions on Medical Imaging* **42**(1), 42–54 (2023) [doi:10.1109/TMI.2022.3203022].
140. C. Han et al., "Meta multi-task nuclei segmentation with fewer training samples," *Medical Image Analysis* **80**, 102481 (2022) [doi:10.1016/j.media.2022.102481].
141. A. Chamanzar and Y. Nie, "Weakly Supervised Multi-Task Learning for Cell Detection and Segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 513–516 (2020) [doi:10.1109/ISBI45749.2020.9098518].
142. C. F. Koyuncu et al., "DeepDistance: A multi-task deep regression model for cell detection in inverted microscopy images," *Medical Image Analysis* **63**, 101720 (2020) [doi:10.1016/j.media.2020.101720].
143. X. Wang et al., "A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images," *Medical Image Analysis* **68**, 101914 (2021) [doi:10.1016/j.media.2020.101914].
144. R. Sarkar et al., *Learning to segment clustered amoeboid cells from brightfield microscopy via multi-task learning with adaptive weight selection*, p. 3852 (2021) [doi:10.1109/ICPR48806.2021.9412641].
145. P. Moeskops et al., "Deep Learning for Multi-task Medical Image Segmentation in Multiple Modalities," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin et al., Eds., pp. 478–486, Springer International Publishing, Cham (2016) [doi:10.1007/978-3-319-46723-8_55].
146. R. A. Caruana, "Multitask Learning: A Knowledge-Based Source of Inductive Bias," 1993, 41–48, Elsevier [doi:10.1016/B978-1-55860-307-3.50012-5].
147. L. Duong et al., "Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 845–850, Association for Computational Linguistics, Beijing, China (2015) [doi:10.3115/v1/P15-2139].
148. A. Kirillov et al., "Segment Anything," arXiv:2304.02643, arXiv (2023) [doi:10.48550/arXiv.2304.02643].
149. A. Vaswani et al., "Attention Is All You Need," arXiv:1706.03762, arXiv (2017) [doi:10.48550/arXiv.1706.03762].
150. R. Bommasani et al., "On the Opportunities and Risks of Foundation Models," arXiv:2108.07258, arXiv (2022) [doi:10.48550/arXiv.2108.07258].
151. I. Lboukili, G. Stamatas, and X. Descombes, "Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using cell morphological features," *J Biomed Opt* **28**(4), 046003 (2023) [doi:10.1117/1.JBO.28.4.046003].
152. I. Lboukili, G. Stamatas, and X. Descombes, "Automatic cell identification and analysis on in vivo reflectance confocal microscopy images of the human epidermis," in *SPIE Photonics Europe 2022*, Strasbourg, France (2022) [doi:10.1117/12.2626777].
153. I. Lboukili, G. N. Stamatas, and X. Descombes, "Age-dependent changes in epidermal architecture explored using an automated image analysis algorithm on in vivo reflectance confocal microscopy images," *Skin Research and Technology* **29**(5), e13343 (2023) [doi:10.1111/srt.13343].
154. N. Din and J. Yu, *Unsupervised deep learning method for cell segmentation* (2021) [doi:10.1101/2021.05.17.444529].
155. V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," 1, *International Journal of Computer Vision* **22**(1), 61–79 (1997) [doi:10.1023/A:1007979827043].

156. P. Marquez-Neila, L. Baumela, and L. Alvarez, "A Morphological Approach to Curvature-Based Evolution of Curves and Surfaces," 1, *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(1), 2–17 (2014) [doi:10.1109/TPAMI.2013.106].
157. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**(6), 610–621 (1973) [doi:10.1109/TSMC.1973.4309314].
158. S. van der Walt et al., "scikit-image: Image processing in Python," preprint, *PeerJ PrePrints* (2014) [doi:10.7287/peerj.preprints.336v2].
159. D. Clausi and Y. Zhao, "Rapid extraction of image texture by co-occurrence using a hybrid data structure," *Computers & Geosciences - COMPUT GEOSCI* **28**, 763–774 (2002) [doi:10.1016/S0098-3004(01)00108-X].
160. Y. Sato et al., "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," 2, *Med Image Anal* **2**(2), 143–168 (1998) [doi:10.1016/S1361-8415(98)80009-1].
161. S. Survarachakan et al., "Effects of Enhancement on Deep Learning Based Hepatic Vessel Segmentation," *Electronics* **10**, 1165 (2021) [doi:10.3390/electronics10101165].
162. J. Lamy et al., "Vesselness Filters: A Survey with Benchmarks Applied to Liver Imaging," in 2020 25th International Conference on Pattern Recognition (ICPR), pp. 3528–3535 (2021) [doi:10.1109/ICPR48806.2021.9412362].
163. T. Barbu, "Gabor filter-based face recognition technique," *Proceedings of the Romanian Academy - Series A: Mathematics, Physics, Technical Sciences, Information Science* **11**, 277–283 (2010).
164. P. K. Saha, G. Borgefors, and G. Sanniti di Baja, "A survey on skeletonization algorithms and their applications," *Pattern Recognition Letters* **76**, 3–12 (2016) [doi:10.1016/j.patrec.2015.04.006].
165. P. Kaur et al., "Hybrid deep learning for Reflectance Confocal Microscopy skin images," in 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 1466–1471 (2016) [doi:10.1109/ICPR.2016.7899844].
166. R. Marceau and Y. Usson, "Methods for the study of cellular sociology: Voronoi diagrams and parametrization of the spatial relationships," 3, *Journal of Theoretical Biology* **154**(3), 359–369 (1992) [doi:10.1016/S0022-5193(05)80176-6].
167. F. Sheikhzadeh et al., "Quantification of confocal fluorescence microscopy for the detection of cervical intraepithelial neoplasia," 1, *BioMedical Engineering OnLine* **14**(1), 96 (2015) [doi:10.1186/s12938-015-0093-6].
168. G. Bigras et al., "Cellular sociology applied to neuroendocrine tumors of the lung: quantitative model of neoplastic architecture," 1, *Cytometry* **24**(1), 74–82 (1996) [doi:10.1002/(SICI)1097-0320(19960501)24:1<74::AID-CYTO9>3.0.CO;2-I].
169. E. Debreuve, "DAccuracy," *GitLab*, <<https://gitlab.inria.fr/edebreuve/daccuracy/-/tree/master/daccuracy>> (accessed 25 January 2022).
170. M. A. Harris et al., "A Pulse Coupled Neural Network Segmentation Algorithm for Reflectance Confocal Images of Epithelial Tissue," 3, *PLoS One* **10**(3), e0122368 (2015) [doi:10.1371/journal.pone.0122368].
171. J. W. Fluhr et al., "Age-Dependent Transformation of Skin Biomechanical Properties and Micromorphology during Infancy and Childhood," *Journal of Investigative Dermatology* **139**(2), 464–466 (2019) [doi:10.1016/j.jid.2018.07.034].
172. G. N. Stamatas et al., "Infant skin microstructure assessed in vivo differs from adult skin in organization and at the cellular level," *Pediatr Dermatol* **27**(2), 125–131 (2010) [doi:10.1111/j.1525-1470.2009.00973.x].
173. A. G. Doukas et al., "Fluorescence excitation spectroscopy for the measurement of epidermal proliferation," *Photochem Photobiol* **74**(1), 96–102 (2001) [doi:10.1562/0031-8655(2001)074<0096:fesftm>2.0.co;2].

174. J. w. Fluhr et al., "Development and organization of human stratum corneum after birth: electron microscopy isotropy score and immunocytochemical corneocyte labelling as epidermal maturation's markers in infancy," *British Journal of Dermatology* **171**(5), 978–986 (2014) [doi:10.1111/bjd.12880].
175. J. W. Fluhr et al., "Functional skin adaptation in infancy - almost complete but not fully competent," *Exp Dermatol* **19**(6), 483–492 (2010) [doi:10.1111/j.1600-0625.2009.01023.x].
176. J. Nikolovski et al., "Barrier function and water-holding and transport properties of infant stratum corneum are different from adult and continue to develop through the first year of life," *J Invest Dermatol* **128**(7), 1728–1736 (2008) [doi:10.1038/sj.jid.5701239].
177. Imane Lboukili, Georgios Stamatas, and Xavier Descombes, "Automating reflectance confocal microscopy image analysis for dermatological research: a review," *Journal of Biomedical Optics* **27**(7), 070902 (2022) [doi:10.1117/1.JBO.27.7.070902].
178. G. Stamatas and N. Kollias, "Infant skin maturation: Structural changes revealed by in vivo reflect," in *Reflectance Confocal Microscopy of Cutaneous Tumors*, 2nd ed. (2017).
179. L. M. Russell, S. Wiedersberg, and M. Begoña Delgado-Charro, "The determination of stratum corneum thickness An alternative approach," *European Journal of Pharmaceutics and Biopharmaceutics* **69**(3), 861–870 (2008) [doi:10.1016/j.ejpb.2008.02.002].
180. R. R. Wickett and M. O. Visscher, "Structure and function of the epidermal barrier," *American Journal of Infection Control* **34**(10, Supplement), S98–S110 (2006) [doi:10.1016/j.ajic.2006.05.295].
181. A. Baroni et al., "Structure and function of the epidermis related to barrier properties," *Clinics in Dermatology* **30**(3), 257–262 (2012) [doi:10.1016/j.clindermatol.2011.08.007].
182. S. Dinas and J. Bañón, "A REVIEW ON DELAUNAY TRIANGULATION WITH APPLICATION ON COMPUTER VISION," *IJCSE - International Journal of Computer Science and Engineering* **3**, 9–18 (2014).
183. F. Darro et al., "Characterization of the differentiation of human colorectal cancer cell lines by means of Voronoi diagrams," *Cytometry* **14**(7), 783–792 (1993) [doi:10.1002/cyto.990140711].
184. D. Zwillinger and S. Kokoska, *CRC standard probability and statistics tables and formulae*, Chapman & Hall/CRC, Boca Raton (2000).
185. C.-N. A et al., "Cooperation and competition in the dynamics of tissue architecture during homeostasis and tumorigenesis," *Seminars in cancer biology* **23**(4), Semin Cancer Biol (2013) [doi:10.1016/j.semcancer.2013.05.009].
186. E. Boireau-Adamezyk, A. Baillet-Guffroy, and G. N. Stamatas, "Age-dependent changes in stratum corneum barrier function," *Skin Res Technol* **20**(4), 409–415 (2014) [doi:10.1111/srt.12132].
187. G. N. Stamatas et al., "Facial skin fluorescence as a marker of the skin's response to chronic environmental insults and its dependence on age," *Br J Dermatol* **154**(1), 125–132 (2006) [doi:10.1111/j.1365-2133.2005.06949.x].
188. K. L. McKinley et al., "Cellular aspect ratio and cell division mechanics underlie the patterning of cell progeny in diverse mammalian epithelia," *eLife* **7**, e36739 [doi:10.7554/eLife.36739].
189. Y. Gu and S. Oliferenko, "Cellular geometry scaling ensures robust division site positioning," *1, Nat Commun* **10**(1), 268, Nature Publishing Group (2019) [doi:10.1038/s41467-018-08218-2].
190. L. Avraham-Chakim et al., "Fluid-Flow Induced Wall Shear Stress and Epithelial Ovarian Cancer Peritoneal Spreading," *PloS one* **8**, e60965 (2013) [doi:10.1371/journal.pone.0060965].
191. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]* (2015).
192. N. U. Din and J. Yu, "Unsupervised deep learning method for cell segmentation," p. 2021.05.17.444529, *bioRxiv* (2021) [doi:10.1101/2021.05.17.444529].
193. O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge" (2014) [doi:10.48550/arXiv.1409.0575].

194. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" (2019) [doi:10.48550/arXiv.1905.11946].
195. C. H. Sudre et al., "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations" (2017) [doi:10.1007/978-3-319-67558-9_28].
196. T.-Y. Lin et al., "Focal Loss for Dense Object Detection" (2017) [doi:10.48550/arXiv.1708.02002].
197. W. D. Cameron et al., "Leveraging multimodal microscopy to optimize deep learning models for cell segmentation," *APL Bioeng* **5**(1), 016101 (2021) [doi:10.1063/5.0027993].
198. J.-Y. Zhu et al., "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" (2017) [doi:10.48550/arXiv.1703.10593].
199. R. Gulakala, B. Markert, and M. Stoffel, "Generative adversarial network based data augmentation for CNN based detection of Covid-19," *1, Sci Rep* **12**(1), 19186, Nature Publishing Group (2022) [doi:10.1038/s41598-022-23692-x].
200. N. Hernandez-Cruz, D. Cato, and J. Favela, "Neural Style Transfer as Data Augmentation for Improving COVID-19 Diagnosis Classification," *SN Computer Science* **2**(5) (2021) [doi:10.1007/s42979-021-00795-2].
201. H. Huang, P. S. Yu, and C. Wang, "An Introduction to Image Synthesis with Generative Adversarial Nets," arXiv:1803.04469, arXiv (2018) [doi:10.48550/arXiv.1803.04469].
202. R. Barth, J. Hemming, and E. J. van Henten, "Improved Part Segmentation Performance by Optimising Realism of Synthetic Images using Cycle Generative Adversarial Networks," arXiv:1803.06301, arXiv (2018) [doi:10.48550/arXiv.1803.06301].
203. Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised Cross-Domain Image Generation" (2016) [doi:10.48550/arXiv.1611.02200].
204. I. J. Goodfellow et al., "Generative Adversarial Networks," arXiv:1406.2661, arXiv [doi:10.48550/arXiv.1406.2661].
205. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, JMLR Workshop and Conference Proceedings (2010).
206. N. Ks, G. D, and R. M, "Skin imaging with reflectance confocal microscopy," *Seminars in cutaneous medicine and surgery* **27**(1), Semin Cutan Med Surg (2008) [doi:10.1016/j.sder.2008.01.006].
207. M. Crawshaw, "Multi-Task Learning with Deep Neural Networks: A Survey" (2020) [doi:10.48550/arXiv.2009.09796].
208. P. Vafaeikia, K. Namdar, and F. Khalvati, "A Brief Review of Deep Multi-task Learning and Auxiliary Task Learning" (2020) [doi:10.48550/arXiv.2007.01126].
209. H. Edelsbrunner, "Smooth surfaces for multi-scale shape representation," in *Foundations of Software Technology and Theoretical Computer Science*, pp. 391–412, Springer, Berlin, Heidelberg (1995) [doi:10.1007/3-540-60692-0_63].
210. U. Schmidt et al., "Cell Detection with Star-convex Polygons," pp. 265–273 (2018) [doi:10.1007/978-3-030-00934-2_30].
211. I. Lboukili, G. Stamatas, and X. Descombes, "Automatic cell identification and analysis on in vivo reflectance confocal microscopy images of the human epidermis," in *SPIE Photonics Europe 2022*, Strasbourg, France (2022) [doi:10.1117/12.2626777].
212. O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *Int J Comput Vis* **115**(3), 211–252 (2015) [doi:10.1007/s11263-015-0816-y].
213. A. Lehmussola et al., "Computational framework for simulating fluorescence microscope images with cell populations," *IEEE Trans Med Imaging* **26**(7), 1010–1016 (2007) [doi:10.1109/TMI.2007.896925].
214. C. Edlund et al., "LIVECell—A large-scale dataset for label-free live cell segmentation," *9, Nat Methods* **18**(9), 1038–1045, Nature Publishing Group (2021) [doi:10.1038/s41592-021-01249-6].

215. K. Clark et al., "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *J Digit Imaging* **26**(6), 1045–1057 (2013) [doi:10.1007/s10278-013-9622-7].
216. S. Roy et al., "A study about color normalization methods for histopathology images," *Micron* **114**, 42–61 (2018) [doi:10.1016/j.micron.2018.07.005].
217. M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1107–1110 (2009) [doi:10.1109/ISBI.2009.5193250].
218. C. A. Barbano and A. Pedersen, "EIDOSLAB/torchstain: v1.2.0-stable," Zenodo (2022) [doi:10.5281/zenodo.6979540].
219. Z. Liu et al., *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows* (2021).
220. S. Schuh et al., "Line-Field Confocal Optical Coherence Tomography: A New Tool for the Differentiation between Nevi and Melanomas?," *Cancers (Basel)* **14**(5), 1140 (2022) [doi:10.3390/cancers14051140].
221. C. Ruini et al., "In-Vivo LC-OCT Evaluation of the Downward Proliferation Pattern of Keratinocytes in Actinic Keratosis in Comparison with Histology: First Impressions from a Pilot Study," *Cancers (Basel)* **13**(12), 2856 (2021) [doi:10.3390/cancers13122856].
222. E. Cinotti et al., "Diagnostic Accuracy of Line-Field Confocal Optical Coherence Tomography for the Diagnosis of Skin Carcinomas," *3, Diagnostics* **13**(3), 361, Multidisciplinary Digital Publishing Institute (2023) [doi:10.3390/diagnostics13030361].
223. L.-K. Huang and M.-J. J. Wang, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recognition* **28**(1), 41–51 (1995) [doi:10.1016/0031-3203(94)E0043-K].
224. J. M. S. Prewitt and M. L. Mendelsohn, "The Analysis of Cell Images*," *Annals of the New York Academy of Sciences* **128**(3), 1035–1053 (1966) [doi:10.1111/j.1749-6632.1965.tb11715.x].
225. "Picture Thresholding Using an Iterative Selection Method," *IEEE Transactions on Systems, Man, and Cybernetics* **8**(8), 630–632 (1978) [doi:10.1109/TSMC.1978.4310039].
226. C. H. Li and P. K. S. Tam, "An iterative algorithm for minimum cross entropy thresholding," *Pattern Recognition Letters* **19**(8), 771–776 (1998) [doi:10.1016/S0167-8655(98)00057-9].
227. J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing* **29**(3), 273–285 (1985) [doi:10.1016/0734-189X(85)90125-2].
228. C. A. Glasbey, "An Analysis of Histogram-Based Thresholding Algorithms," *CVGIP: Graphical Models and Image Processing* **55**(6), 532–537 (1993) [doi:10.1006/cgip.1993.1040].
229. J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition* **19**(1), 41–47 (1986) [doi:10.1016/0031-3203(86)90030-0].
230. W.-H. Tsai, "Moment-preserving thresholding: A new approach," *Computer Vision, Graphics, and Image Processing* **29**(3), 377–393 (1985) [doi:10.1016/0734-189X(85)90133-1].
231. W. Doyle, "Operations Useful for Similarity-Invariant Pattern Recognition," *J. ACM* **9**(2), 259–267 (1962) [doi:10.1145/321119.321123].
232. A. G. Shanbhag, "Utilization of Information Measure as a Means of Image Thresholding," *CVGIP: Graphical Models and Image Processing* **56**(5), 414–419 (1994) [doi:10.1006/cgip.1994.1037].
233. G. W. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency," *J Histochem Cytochem.* **25**(7), 741–753, *Journal of Histochemistry & Cytochemistry* (1977) [doi:10.1177/25.7.70454].
234. J.-C. Yen, F.-J. Chang, and S. Chang, "A new criterion for automatic multilevel thresholding," *IEEE Transactions on Image Processing* **4**(3), 370–378 (1995) [doi:10.1109/83.366472].
235. J. Li et al., "Biopsy-free in vivo virtual histology of skin using deep learning," *1, Light Sci Appl* **10**(1), 233, Nature Publishing Group (2021) [doi:10.1038/s41377-021-00674-8].

Appendix

Gabor filter bank

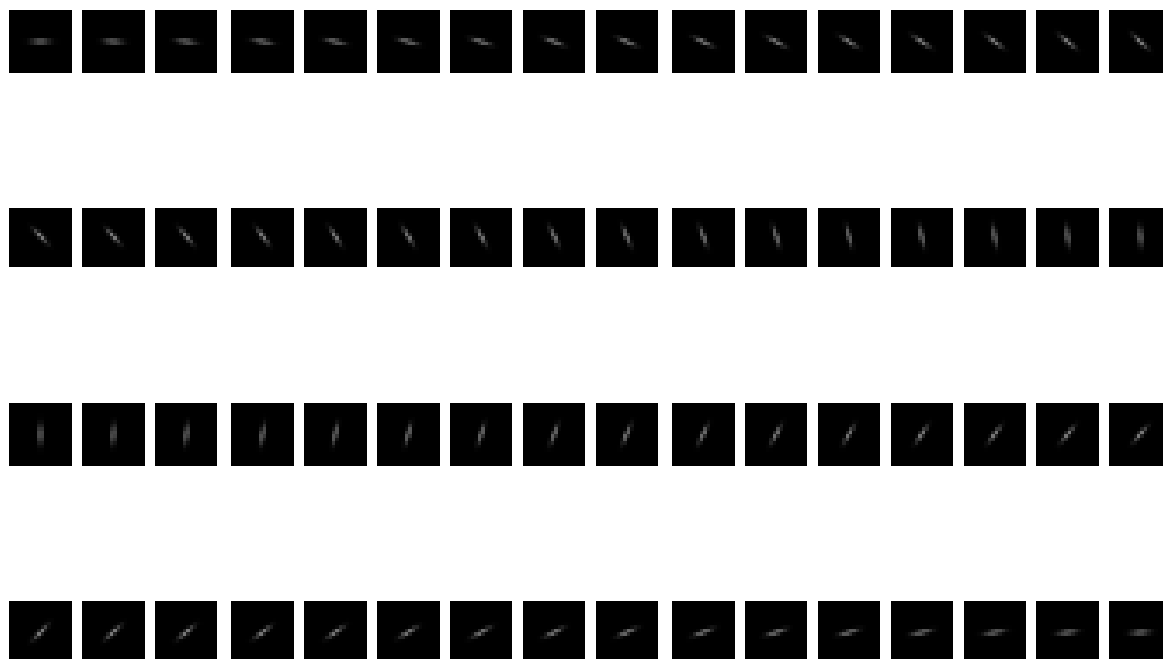
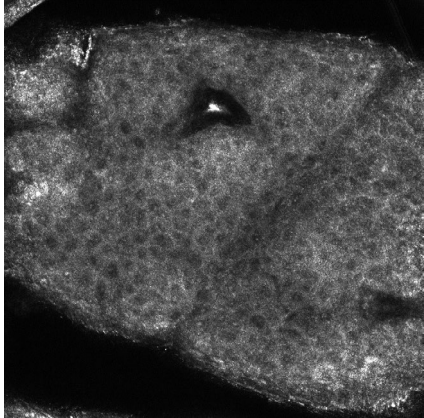
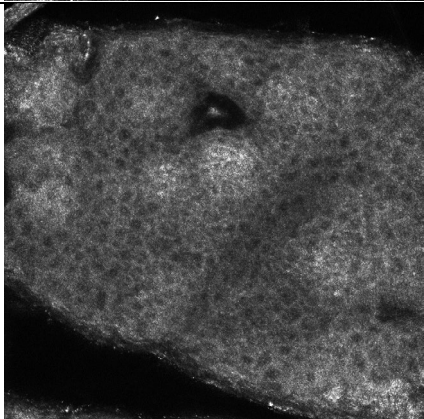


Fig. S1 Gabor filters applied to the RCM image during the FIAP. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

Application of the FIAP to RCM images of dark-pigmented skin

Table S1 Cell detection accuracy on 2 *stratum granulosum* RCM images of dark-pigmented skin for Expert 1 using FIAP. RCM, Reflectance confocal microscopy; FIAP, Full image analysis pipeline.

Image	Layer	Precision (%)	Recall (%)	F1-score (%)
	SG	62.3	70.2	66.0
	SG	68.3	66.7	67.5

DermoGAN

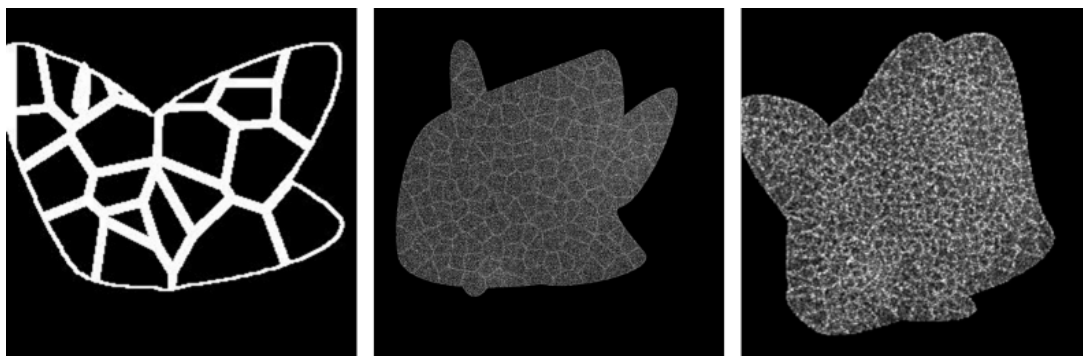


Fig. S2 Synthetic images used in the training of the models. On the left, a synthetic binary image used in the DermoGAN training, and on the right synthetic RCM images of different noise levels and cell sizes used in U-net training. RCM, Reflectance confocal microscopy.

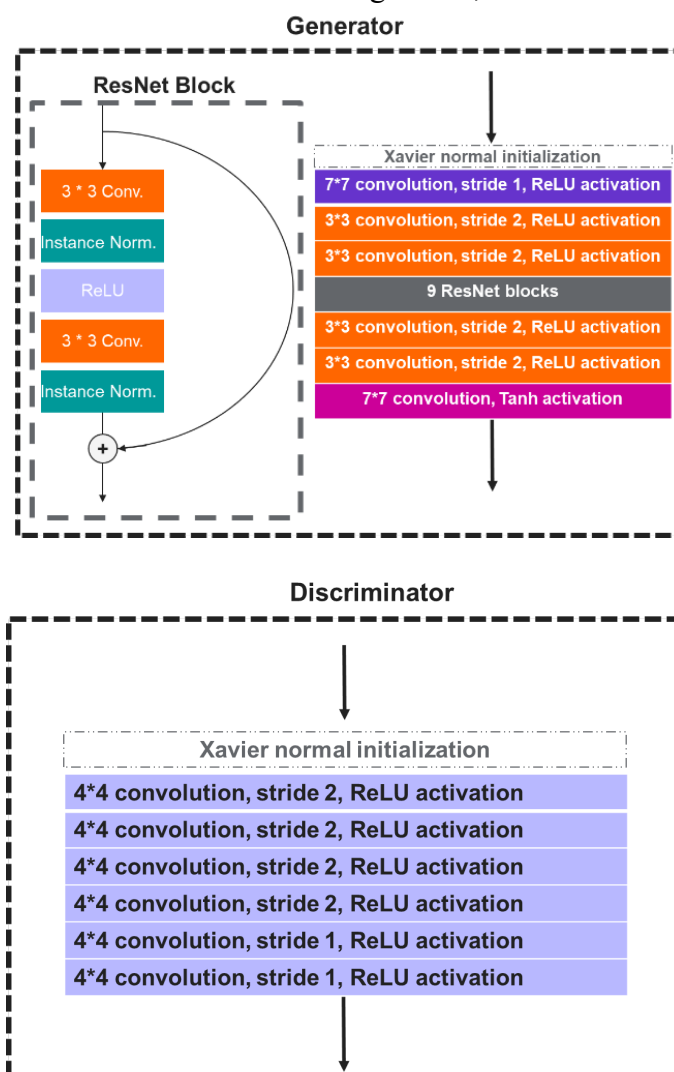


Fig. S3 Structure of the generator and discriminator networks used in the cycle-GAN and DermoGAN approaches. RCM, Reflectance confocal microscopy.

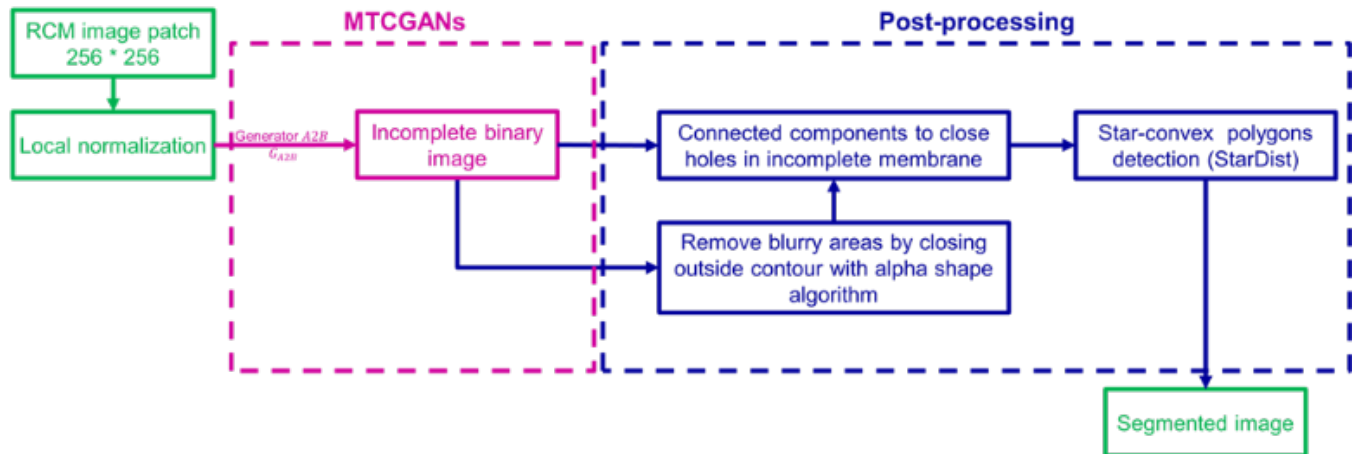


Fig. S4 To obtain keratinocytes positions. We apply the $G_{A2B}: A \rightarrow B$ network to locally normalized RCM image and obtain an incomplete cell identification, which is then cleaned by closing any holes in the detected membrane and the outside contour, and finally the cell identification is refined using StarDist algorithm. RCM, reflectance confocal microscopy.

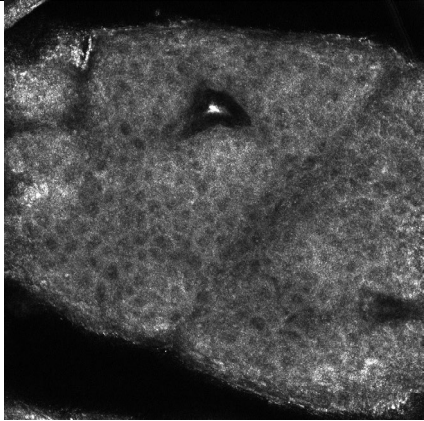
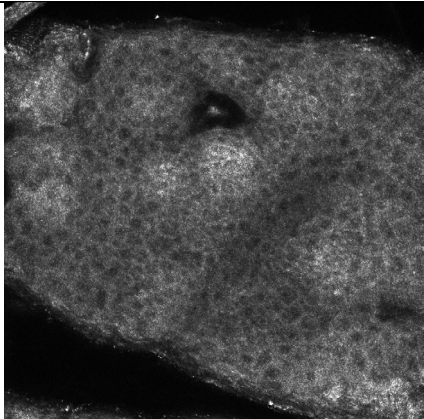
Accuracy metrics of the application of DermoGAN compared to those of 5 other approaches

Table S2 Comparison of all accuracy metrics for all six tested approaches (in %).

Image	Layer	U-net pre-trained and augmented with real and synthetic images with Focal and Dice loss functions			StarDist on Gabor-filtered images			Cycle-GAN using RCM images			Cycle-GAN using Gabor-filtered images			Traditional approach			DermoGAN		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
1	SS	50,3	29,8	37,5	37,1	47,5	41,6	47,8	46,8	47,3	42,9	40,8	41,8	46,2	89,4	61	68,4	73,6	70,9
2	SS	37,4	24,6	29,7	50,3	53,8	51,9	40,6	51,7	45,5	57,7	33,9	42,7	52	76,1	61,8	77,9	51	61,6
3	SS	40,8	24,6	30,7	50,5	44,4	47,3	45,9	42,6	44,2	50,4	26,7	34,9	48,9	79,6	60,6	72,5	60,7	66,1
4	SG	44,8	51,3	47,9	43,8	76,1	55,6	31,9	65,2	42,9	30,6	41,3	35,2	63,9	60,5	62,2	64,9	82	72,5
5	SG	28,8	56,8	38,2	27,7	65,5	38,9	23,1	67,3	34,4	17,5	32,7	22,8	56,4	77	65,1	53,4	85,1	65,6
6	SG	45,3	64,3	53,2	14,4	72	24	29,9	68,4	41,6	22,2	28	24,8	71,3	81,8	76,2	53,2	84,8	65,4
7	SG	42,4	70,7	53	14,7	78,7	24,7	19,7	58,7	29,5	16,8	31,8	22	71,6	90,7	80	63,9	92	75,4
8	SG	60,5	55,3	57,8	23	62,8	33,7	30,1	43,6	35,7	34,4	27,3	30,4	78,9	56,3	65,7	65,8	75,8	70,4
9	SG	50,5	86,6	63,8	14,8	79,5	25	21	70,9	32,4	18,8	26,1	21,8	64,1	87,4	74	57,4	86,6	69

Application of DermoGAN to RCM images of dark-pigmented skin

Table S3 Cell detection accuracy on 2 *stratum granulosum* RCM images of dark-pigmented skin for Expert 1 using DermoGAN.

Image	Layer	Precision (%)	Recall (%)	F1-score (%)
	SG	0.365 ± 0.15	0.803 ± 0.14	0.502 ± 0.16
	SG	0.487 ± 0.17	0.769 ± 0.11	0.596 ± 0.14

Thresholding approaches in ImageJ used in the evaluation of the accuracy of DermoGAN applied to fluorescence microscopy images

1. Default thresholding in ImageJ: This is the original method of auto thresholding available in ImageJ, which is a variation of the IsoData algorithm.
2. Huang's fuzzy thresholding method²²³.
3. Huang2 thresholding: This is an alternative implementation of Huang's method.
4. Intermodos thresholding²²⁴.
5. IsoData thresholding²²⁵.
6. Li's Minimum Cross Entropy thresholding²²⁶ method based on the iterative version of the algorithm.
7. MaxEntropy thresholding: Implements Kapur-Sahoo-Wong²²⁷ (Maximum Entropy) thresholding method.
8. Mean thresholding²²⁸.
9. MinError(I): An iterative implementation of Kittler and Illingworth's Minimum Error thresholding²²⁹.
10. Minimum thresholding²²⁴.
11. Moments: Implements Tsai's method²³⁰ for thresholding.
12. Otsu thresholding⁶⁸.
13. Percentile thresholding²³¹:
14. Renyi Entropy: Similar to the MaxEntropy method, but using Renyi's entropy instead.
15. Shanbhag thresholding²³².
16. Triangle thresholding²³³.
17. Yen's thresholding²³⁴.

Transforming RCM into histology and *vice versa*

Throughout this manuscript, we have mentioned the advantages of RCM images in comparison to histology, from being used to study healthy (baby) skin, to reducing the number of unnecessary biopsies by 2-3 times and providing information for the diagnosis and monitoring of multiple diseases. Unfortunately, RCM use is limited by the complexity of extraction of information from the generated images and by the limited number of dermatologists trained to read RCM images vs. histopathology slides. Therefore, developing a method to leverage the insights provided by RCM through the lens of histology would be useful in clinical practice. We compare a cycle-GAN algorithm to the developed DermoGAN for the virtual staining of RCM images.

Similar work has been done by Li *et al.*²³⁵ using *ex vivo* RCM images, of normal skin, skin with basal cell carcinoma, and skin presenting melanocytic nevi, for the training of the proposed deep learning model. Indeed, a GAN was trained using RCM images of excised skin with and without acetic acid nuclear contrast staining, which provides nuclear contrast in RCM images, to learn the paired translation between stained and unstained RCM images. The resulting stained images (virtually or not) of both *in vivo* and *ex vivo* RCM images were then virtually stained with hematoxylin and eosin to resemble histology images.

Our approaches will be limited to RCM images of healthy skin and serves as an exploration of what can be done with the proposed models.

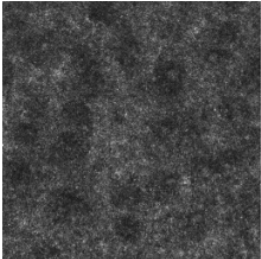
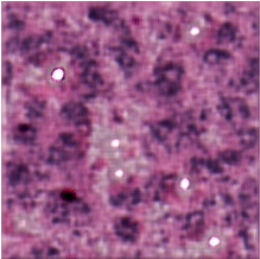
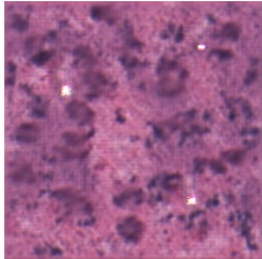
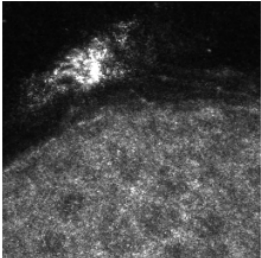
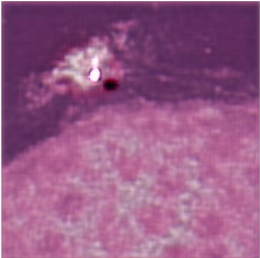
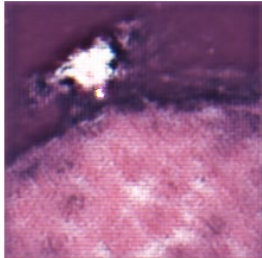
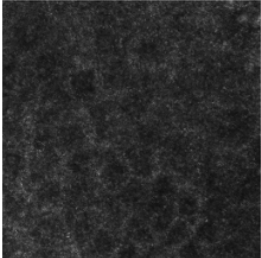
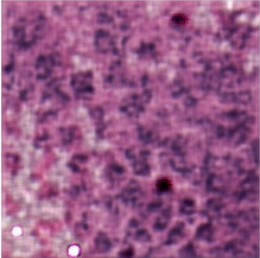
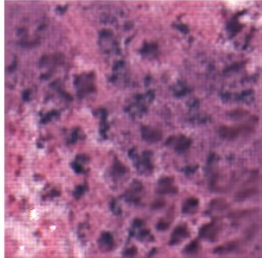
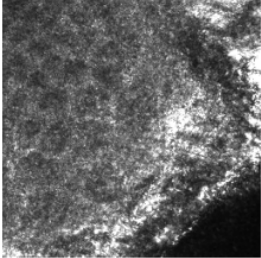
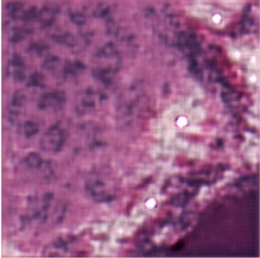
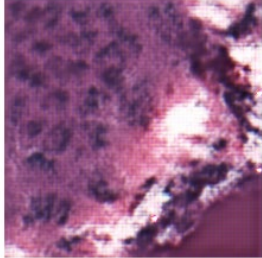
The first tested approach for unpaired RCM to histology translation, and *vice versa*, was based on a cycle-GAN, where the first domain was RCM images of the epidermis, while the second was of histology slides of the skin.

The second tested approach was based on a DermoGAN, where the first task was made of the previously described cycle-GAN, while the second task was a mapping between Gabor-filtered RCM images and histology images.

In both methods, images were processed to represent only the epidermis and be of size 256x256 pixels.

Results of these approaches are far from perfect, but the DermoGAN approach seems to lead to more cohesive nuclei compared to the cycle-GAN approach. Testing DermoGAN with the added intermediate step with acetic acid nuclear contrast staining would be of interest. Additionally, the approach is limited by the available data and presents a major issue of directionality. Indeed, the orientation of RCM images is perpendicular to the vertical sections typical in histopathology, and therefore the attempt at translating RCM images into histology is based on quite different data sets, calling for the question of the feasibility of this approach without acquiring horizontal histology slides, or reconstructing the RCM stack. Both solutions have their limitations, the first is hindered by the need for a new clinical study, while the second is limited by the technical difficulty of obtaining 3D-RCM stacks, made more difficult by the significant difference in resolution between the two methods, with RCM having a horizontal resolution of $0.5 - 1 \mu\text{m}$ and a vertical resolution (optical section thickness) of $3 - 5 \mu\text{m}$, and histology having a resolution of $0.25 - 0.5 \mu\text{m/pixel}$.

Table S4 RCM to histology translation results for a trained cycle-GAN and a trained DermoGAN.

RCM image	Histology image generated with cycle-GAN	Histology image generated with DermoGAN
		
		
		
		

Publications

Journals

1. **Lboukili** et al., *Automating reflectance confocal microscopy image analysis for dermatological research: a review*. Journal of Biomedical Optics 2022 Jul;27(7).
2. **Lboukili** et al., *Automatic granular and spinous epidermal cell identification and analysis on in vivo reflectance confocal microscopy images using cell morphological features*. Journal of Biomedical Optics 2023 Apr;28(4).
3. **Lboukili** et al., *Age-dependent changes in epidermal architecture explored using an automated image analysis algorithm on in vivo reflectance confocal microscopy image*. Skin Research & Technology 2023 May;29(5).
4. (Unrelated to Ph.D. project, but uses data from same clinical studies) Stamatas, Roux, Boireau-Adamezyk, **Lboukili**, Oddos, Skin maturation from birth to 10 years of age: Structure, function, composition and microbiome. Experimental dermatology 2023 June.
5. (Unrelated to Ph.D. project, but uses data from same clinical studies) Sato, Nikolovski, Gould, **Lboukili**, *Differential expression of non-invasive biomarkers associated with risk susceptibility to atopic dermatitis in children with family history of allergic disease* – Submitted to Journal of the European Academy of Dermatology and Venereology

Conference proceedings

Lboukili et al., *Automatic cell identification and analysis on in vivo reflectance confocal microscopy images of the human epidermis* ». In SPIE Photonics Europe 2022. Strasbourg, France, 2022.

Conference talks

SPIE Photonics Europe - April 2022 (Strasbourg, France)	Automatic cell identification and analysis on in vivo reflectance confocal images of the human epidermis
--	--

Conference talks

- | | |
|--|---|
| SPIE Photonics Europe - April 2022
(Strasbourg, France) | <ul style="list-style-type: none">- Automated identification of keratinocytes on in vivo reflectance confocal microscopy images of the human skin epidermis- (Unrelated to PhD, but uses data from same clinical studies) Effect of hormonal status on skin physiology in postmenopausal women |
| ESDR – September 2022 (Amsterdam, Netherlands) | Age-dependent changes in geometric and topologic characteristics of epidermal keratinocytes using an automated image analysis algorithm on in vivo reflectance confocal microscopy images |
| World congress on confocal microscopy | Methods for automatic cell segmentation of in vivo reflectance confocal microscopy images of the human epidermis |

Patent Applications

1. Patent 63/244,981: Analysis and characterization of epithelial tissue structure
2. (Unrelated to Ph.D. project, but uses data from same clinical studies) Patent 63/304,290: Biomarkers predictive of atopic dermatitis
3. Patent 63/447,374: Systems and methods for automatic cell identification using images of human epithelial tissue structure