UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Ecole Doctorale IAEM-Lorraine
Centre de recherche en automatique de Nancy

Laboratoire de biophotonique
Université d'État de Tomsk (Cotutelle)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention : "Automatique, Traitement du signal et des images, Génie informatique"

par **Valentin Kupriyanov**

Combination of multimodal optical spectroscopy and machine learning methods for non invasive in vivo diagnostic aid of precancerous and cancerous skin lesions in preclinical and clinical studies

Presented and publicly defended on November 13th, 2024

Membres du jury :

Rapporteurs :	**Anne HUMEAU-HEURTIER**	Professeur, Laboratoire Angevin de Recherche en Ingénierie des Systèmes, Université d'Angers, Angers, France
	Andrey DUNAEV	Chercheur principal, Université d'État de Orel, Orel, Russie
Examinateurs :	**Olga M. CONDE PORTILLA**	Professeure, Université de Cantabrie, Cantabrie, Espagne
	Martin HOHMANN	Doktoringenieur, Institute of Photonic Technologies, Erlangen, Germany
	Yury KISTENEV	Professeur, Université d'État de Tomsk, Tomsk, Russie Directeur de thèse
	Christian DAUL	Professeur, Centre de recherche en automatique de Nancy, Université de Lorraine, Nancy, France Co-directeur de thèse
	Marine AMOUROUX	Ingénieure de Recherche, Centre de recherche en automatique de Nancy, Université de Lorraine, Nancy, France Co-encadrante de thèse
Président de jury :	**Yann GUERMEUR**	Directeur de recherche, Laboratoire lorrain de recherche en informatique et ses applications, Université de Lorraine, Nancy, France
Invité:	**Walter BLONDEL**	Professeur, Centre de recherche en automatique de Nancy, Université de Lorraine, Nancy, France Directeur de thèse
	Victor CHEREPANOV	Professeur, Université d'État de Tomsk, Tomsk, Russie

Remerciements/Acknowledgements

The preparation of this thesis would not have been possible without the help of a large number of people and organizations, and I am very grateful to all those who have helped me directly or indirectly throughout this time.

In particular, I would like to thank:

To my supervisor at the Centre de Recherche sur l'Automatique (CRAN) and the University of Lorraine (UL), Prof. Walter Blondel, for his comprehensive help and advice to me on both scientific and personal matters, as well as for his optimism in all situations. I would also like to thank him specifically for giving me the opportunity to work at CRAN, to participate in several scientific and academic projects and events, and for the influence he has had on my development as a scientist.

My co-supervisor at CRAN and UL, Dr. Marine Amouroux, for her kindness and support in difficult situations, for her hospitality, for organizing joint events that made me feel at home when I was actually thousands of kilometres away. I would also like to thank her and Dr. Grégoire Khairallah for handling the measurement of all the experimental data used in this thesis.

My co-supervisor at CRAN and UL, Prof. Christian Daul, for his attention to detail, which I hope I have been able to adopt, his willingness to help me correct errors, and his high expertise in data processing. I would also like to thank him for sharing his vast knowledge of wine consumption with me.

My supervisor at the Laser Molecular Imaging and Machine Learning laboratory (LMIML) and Tomsk State University (TSU), Prof. Yury Kistenev, for giving me the opportunity to start my journey in science that eventually brought me to where I am, for his patience and for his approach to working with scientific data that he taught me.

The members of my thesis follow up committee: Prof. Olga Conde, Prof. Victor Cherepanov, and Dr. Yann Guermeur, for their time spent reading my reports and listening to my presentations, as well as for their advice on how to improve the quality of my research. I would like to specifically thank Dr. Yann Guermeur for his incredibly helpful advice and the time he took to share with me his deep knowledge of statistics and multimodal data analysis.

The members of my thesis jury, Prof. Anne Humeau-Heurtier and Prof. Andrey Dunayev, for taking the time to read my thesis and participate in my defence as recognized experts in their respective fields of research.

To my friends and colleagues at CRAN, including Victor Colas, Maria Borozdova, Thomas Elsen, Sergey Zaytsev, Matthieu Renard, Clarice Perrin-Mozet, Marie Camonin, Clement Fauvel, Ahmed Zghal, Yann Bailly, Lucie Lerouge and others for their friendship and support, various joint activities and their willingness to help in any situation.

To my colleagues at LMIML and especially Alexey Borisov and Denis Vrazhnov, for their advice in data processing and analysis.

To the administrative and management staff at UL and TSU for their assistance with the preparation of many formal documents and their continued willingness to help in response to any of my requests for assistance.

Also, I would like to thank the French Embassy in Russia and Campus France for providing me with a scholarship under the Vernadsky Grant, which enabled me to enroll in a PhD program at the University of Lorraine and begin my research in France.

I am also very grateful to the University of Lorraine, which kindly agreed to supplement the Vernadsky grant so that I would not have to interrupt my research.

I am also very grateful to the University of Lorraine for the opportunity to participate in the DrEAM academic mobility program and to the Institute for Systems and Computer Engineering, Technology and Science (Porto, Portugal) for hosting me for this project.

Finally, I apologize to everyone I forgot to mention.

Table of contents

Remerciements/Acknowledgements **2**

List of acronyms **8**

List of Tables **11**

List of Figures **13**

Résumé **16**

General Introduction **21**

The list of scientific contributions **25**

Chapter 1: Review of existing methods for the diagnosis of non-melanoma skin cancer **28**

1.1. Introduction 28

1.2. Clinical methods of skin cancer diagnosis 28

1.2.1. Skin structure and composition 28

1.2.1.1. Epidermis 29

1.2.1.2. Dermis and Hypodermis 30

1.2.2. Skin cancers 31

1.2.2.1. Basal Cell Carcinoma 32

1.2.2.2. Squamous-Cell Carcinoma 33

1.2.2.3. Actinic Keratoses 34

1.2.3. Standard clinical procedure for skin cancer diagnosis and treatment 35

1.2.3.1.Dermoscopy 36

1.2.3.2. Biopsy and/or surgical treatment 37

1.2.3.3. Histological analysis 39

1.2.4. Present limitations and needs for clinicians 39

1.3. *In vivo* skin Optical biopsy 40

1.3.1. Changes in skin optical properties related to the development of carcinomas 40

1.3.2. Single modality studies 43

1.3.2.1. Spectroscopic methods 43
1.3.2.2. Imaging methods 50
1.3.3. Multimodal approaches 55
1.4. Conclusion 60
Chapter 2. Spectroscopic data processing and analysis methods 61
2.1. Introduction 61
2.2. Pre-processing of raw data 62
2.2.1. Noise removal 62
2.2.2. Outlier removing 63
2.2.3. Normalization, Standardization and Scaling 65
2.2.4. Dimensionality reduction 66
2.2.4.1. Feature selection methods 67
2.2.4.2. Feature extraction methods 68
2.2.5. The use of machine learning methods at pre-processing stage 69
2.3. Data processing 70
2.3.1. Supervised Classification 71
2.3.2. Regression 77
2.3.3. Unsupervised Classification (Clustering) 80
2.4. Data fusion 81
2.5. Quantitative evaluation of the performance of data analysis algorithms 82
2.5.1. Statistical metrics for classification 82
2.5.2. Accuracy metrics for regression 86
2.5.3. Splitting of the data set 87
2.6. Conclusion 90
Chapter 3. Pre-clinical study on mice skin 91
3.1. Introduction 91

3.2. Materials and methods 92
3.2.1. Characteristics of the experimental device 92
3.2.2. Characteristics of the dataset under study 93
3.2.3. Spectrum analysis pipeline 99
3.2.3.1. Pre-processing of the spectra 99
3.2.3.2 Classification of the spectra 100
3.2.3.3 Data Fusion 101
3.3. Results and Discussion 103
3.3.1. Results of methods and hyperparameters selection 103
3.3.2. Results of classification for different groups of spectra 106
3.3.3. Results of Data fusion 110
3.4. Conclusion 113
Chapter 4. Clinical study on human skin 115
4.1. Introduction 115
4.2. Materials and methods 115
4.2.1. General information about clinical study 115
4.2.2. Instrumentation and measurement procedure 116
4.2.3. Clinical and spectroscopic dataset 118
4.2.4. Spectroscopic data processing pipeline 124
4.2.4.1. Raw spectra filtering and correction 124
4.2.4.2. Detection of outliers in series of measurements 125
4.2.4.3. Normalization of spectra 127
4.2.4.4. Dimensionality reduction 128
4.2.4.5. Classification 129
4.2.5 Data Fusion 130
4.3 Results and Discussion 131

4.3.1. Selection of best methods and hyperparameters for the data processing pipeline 131
4.3.2. Binary classification results 132
4.3.2.1 BCC, SCC, AK and H classes 132
4.3.2.2 CI, CNI, AK and H classes 134
4.3.3. Multiclass classification results 136
4.3.4. Evaluation of the results obtained 138
4.3.4.1. In relation to the preclinical study 138
4.3.4.2. In relation to other studies on the use of AF and DR spectra for skin cancer diagnosis 139
4.3.5. The use of other data pre-processing and processing methods 139
4.4. Conclusion 141
General Conclusion and Prospects 143
General Conclusion 143
Prospects 144
List of references 145

List of acronyms

Acc – Accuracy

AE - AutoEncoder

AF – Auto-Fluorescence

AFS – Auto-Fluorescence Spectroscopy

AH – Atypical Hyperplasia

AK – Actinic Keratosis

ANN – Artificial Neural Network

AUC – Area Under Curve

BCC – Basal Cell Carcinoma

CH – Compensatory Hyperplasia

CI – Carcinomas Invasive

CNI – Carcinomas Non-Invasive

CV – Cross-Validation

D – Dysplasia/Dysplastic hyperplasia

DA – Discriminant Analysis

DL – Deep Learning

DR – Diffuse Reflectance

DRS – Diffuse Reflectance Spectroscopy

DT – Decision Tree

FD-OCT – Frequence-Domain OCT

FF-OCT – Full-Field Optical Coherence Tomography

FLIM – Fluorescence Lifetime Imaging Microscopy

FN – False Negative

FP – False Positive

FPR – False Positive Rate

FWHM – Full Width at Half Maximum

H – Healthy/Healthy skin

HSI – Hyper-Spectral Imaging

ICA – Independent Component Analysis

IR – Infra-Red

LC-OCT – Line-field Confocal Optical Coherence Tomography

L – Lesion/Lesional

LDA – Linear Discriminant Analysis

LED – Light-Emitting Diode

LOOCV – Leave-One-Out Cross Validation

LPOCV – Leave-*p*-Out Cross Validation

LR – Logistic Regression

MAE – Mean Absolute Error

ML – Machine Learning

MLP – MultiLayer Perceptron

MSC – Melanoma Skin Cancer

MSE – Mean Squared Error

MSI – Multi-Spectral Imaging

MPM – Multi-Photon Microscopy

NIR – Near Infra-Red

NL – Non Lesion/Non-Lesional

NMSC – Non-Melanoma Skin Cancer

OCT – Optical Coherence Tomography

OVO – One-Vs-One

OVA – One-Vs-All

PCA – Principal Component Analysis

PL – Peri-Lesion/Peri-Lesional

QDA – Quadratic Discriminant Analysis

PP – Potentially Precancerous

RBF – Radial Based Function

RCM – Reflectance Confocal Microscopy

RF – Random Forest

ROC – Receiver Operating Characteristic

RS – Raman Spectroscopy

SC – Stratum Corneum

SCC – Squamous Cell Carcinoma

SDS – Source-Detector Separation

Se – Sensitivity

SHG – Second Harmonic Generation

SNR – Signal to Noise Ratio

Sp – Specificity

SVM – Support Vector Machine

TD-OCT – Time-Domain Optical Coherence Tomography

TN – True Negative

TP – True Positive

TPR – True Positive Rate

t-SNE – t-distributed Stochastic Neighbor Embedding

UMAP – Uniform Multiverse Approximation and Projection method

UV – UltraViolet

VC – Variation Coefficient

List of Tables

Chapter 1

Table 1.1. Parameters used for NMSC diagnosis 36

Table 1.2. Summary of the results of the search for articles on NMSC diagnostics using spectroscopic methods 49

Table 1.3. Summary of the results of the search for articles on NMSC diagnostics using imaging methods 55

Table 1.4. Summary of the results of the search for articles on NMSC diagnostics using multimodal methods 58

Chapter 2

-

Chapter 3

Table 3.1. Main light source and detection characteristics of the device used during the experimental study 93

Table 3.2. Main characteristics of the initial set of spectra obtained from mice skin 94

Table 3.3. Symptoms of the skin states under study and their effects on mice skin properties 95

Table 3.4. List of classification methods and corresponding hyperparameters used in preclinical study 101

Table 3.5. Selection of hyperparameters for methods used in preclinical study 106

Table 3.6. Values of accuracy and misclassification for PP states for all modalities and sub-modalities 107

Table 3.7. Confusion matrices corresponding to the most accurate results for every SDS 108

Table 3.8. Results of mice skin spectra binary classification obtained using DR spectra for every SDS 109

Table 3.9. Best results obtained by using fusion of initial data and fusion of features for modalities/sub-modalities 109

Table 3.10. Results obtained by using decision fusion for modalities/sub-modalities 111

Table 3.11: Results of decision fusion by SDS 111

Table 3.12: Classification results based on decisions obtained from all types of spectra 113

Table 3.13: Results of decision fusion for binary classification 113

Chapter 4

Table 4.1. Main light source and detection characteristics of the SpectroLive device 117

Table 4.2. Distribution of the 131 patients in classes of gender, age, phototype and lesion anatomical site 118

Table 4.3. Wavelength ranges [λ_1:λ_2] for the calculation of the average variation coefficients of AF spectra 126

Table 4.4. Data analysis methods and optimized hyperparameters 130

Table 4.5. Best values of hyperparameters retained for each method in the spectroscopic data classification pipeline 131

Table 4.6. Results of decision fusion for BCC, SCC, AK and H classes 134

Table 4.7. Results of decision fusion for CI, CNI, AK and H classes 136

Table 4.8. Best results of single modality multiclass classification by main diagnostic classes 138

List of Figures

Chapter 1

Figure 1.1. Scheme of the human skin layers 29

Figure 1.2. Scheme of the epidermis sub-layers in human skin 29

Figure 1.3. Scheme of the dermis sub-layers and hypodermis structure in human skin 31

Figure 1.4. The principle of Punch Biopsy 38

Figure 1.5. Absorption index and transport extinction index of melanin, blood, basic tissue, oxyhemoglobin, and deoxyhemoglobin 42

Figure 1.6. Example of AF spectra of BCC, SCC, heathy skin and benign lesion 44

Figure 1.7. Example of DR spectra of cancerous and benign tissues 47

Figure 1.8. Example of Raman spectra of cutaneous lesions of different types 48

Figure 1.9. Images of human skin obtained from BCC lesion and healthy skin using Reflectance Confocal Microscopy (A), Fluoresce Lifetime Imaging Microscopy (B) Multiphoton Microscopy and Optical Coherence Tomography (D) 50

Figure 1.10. Images of human skin obtained from MM and using Multispectral Imaging 51

Figure 1.11. Difference between RGB imaging, multispectral imaging and hyperspectral imaging 54

Chapter 2

Figure 2.1. Example of spectra before (a) and after (b) removing of noise 63

Figure 2.2. Collective and Point outliers 64

Figure 2.3. Example of normalisation to intensity ratio for AF and DR spectra 66

Figure 2.4. Illustration representing two first principal components for a two-dimensional original feature space 68

Figure 2.5. Main structure of Autoencoder with one hidden layer 70

Figure 2.6. Example of using the kernel trick with polynomial kernel 72

Figure 2.7. Example of Binary Decision Tree implementation for a dataset with six classes 73

Figure 2.8. Example of a confusion matrix 82

Figure 2.9. Example of a ROC-curve 85

Figure 2.10. Example of *k*-fold cross validation for $k = 5$ 89

Chapter 3

Figure 3.1. Scheme of the pilot device used in preclinical study 92

Figure 3.2. Mean spectra for all classes obtained at SDS_1 = 271μm 96
Figure 3.3. Mean spectra for all classes obtained at SDS_2 = 536μm 97
Figure 3.4. Mean spectra for all classes obtained at SDS_3 = 834μm 99
Figure 3.5. The main scheme of spectra pre-processing 100
Figure 3.6. The main scheme of Data Analysis by using weighted voting 102
Figure 3.7. The main scheme of Data Analysis by using Stacking 103
Figure 3.8. 3D distribution of the first three principal components for all four classes of spectra 104
Figure 3.9. Dependence of the classification accuracy on the number of principal components used 104
Figure 3.10. Average classification accuracy obtained by different classifiers (SVM, MLP, DA and RBF) for all modalities (DR and all AF) and all SDS; error bars indicate standard deviation (SD) values corresponding to five repeated calculations for each combination of source, detector and classifier 107
Figure 3.11. Average accuracy for stacking (decision fusion) obtained by different classifiers (SVM, MLP, DA and RBF) for all ways to combine data (by SDS, by modality and all data at once); error bars indicate SD values corresponding to five repeated calculations for each combination of source, detector and classifier 112

Chapter 4

Figure. 4.1 Simplified schematic representation of the SpectroLive medical device including a Xenon lamp wide-band source for DR measurements and five LED narrow-band sources for AF excitations, the multiple-fiber optics probe featuring 4 Source-Detector Separations (SDS1-4), a filter wheel equipped with long-pass emission filters for AF measurements and 4 spectrometers 116
Figure 4.2. Schematic representation of the spectroscopic measurement spots on: (left) Lesion (L), Peri-Lesion (PL) and Non Lesion (NL) areas in and around the lesion area (red circle) and along the short and long axes (blue lines) of the blue ink delimited surgical spindle; (right) healthy areas on palmar hand and inner wrist 119
Figure 4.3. Mean spectra for main diagnostic classes obtained at SDS_1 = 400μm 120
Figure 4.4. Mean spectra for main diagnostic classes obtained at SDS_2 = 600μm 121
Figure 4.5. Mean spectra for main diagnostic classes obtained at SDS_3 = 800μm 122
Figure 4.6. Mean spectra for main diagnostic classes obtained at SDS_4 = 1000μm 123
Figure 4.7. Spectroscopic data processing pipeline diagram overview 124

Figure 4.8. General scheme of filtering of outliers in measurements 127
Figure 4.9. General scheme of fusion of decisions 130
Figure 4.10 Results of pairwise binary classification for BCC, SCC, AK and H classes, for every modality (AF1..5, DR) and SDS 132
Figure 4.11. Results of classification for dividing cancer into invasive and non-invasive types for every modality (AF1..5, DR) and SDS 135
Figure 4.12. Results of multiclass classifications for every modality (AF1..5, DR) and SDS 137

Résumé

Introduction et définition des objectifs de thèse

En 2022, le carcinome cutané était le cinquième type de cancer le plus fréquent dans le monde [1]. Le carcinome basocellulaire et le carcinome spinocellulaire sont les types de CSNM les plus courants, représentant respectivement 70 % et 25 % des cas. Les métastases ne surviennent que dans un cas sur 14 000 000 pour le carcinome basocellulaire et dans moins de 10 % pour le carcinome spinocellulaire [2], ce qui se traduit par un taux de survie global de 90 % à cinq ans. Le taux d'incidence normalisé selon l'âge du NMSC est passé de 54,08/100 000 en 1990 à 79,10/100 000 en 2019 (+1,78 % par an) et le nombre de cas déclarés augmente chaque année [3–5]. Cette augmentation de l'incidence entraîne une charge médico-économique importante.

En effet, la procédure standard de diagnostic des lésions cutanées suspectes comprend une biopsie chirurgicale et une analyse histopathologique des échantillons de tissus prélevés [6, 7]. Ainsi, selon [8], le ratio biopsies/traitement est passé de 1,1 en 1993 à 2,1 en 2016. En outre, la biopsie est une procédure invasive, longue et coûteuse, qui induit du stress et des cicatrices, et dont la précision diagnostique est faible, puisque seulement 20 % des biopsies révèlent des cancers de la peau et que deux tiers des carcinomes ne sont pas diagnostiqués.

Un autre aspect important est l'évaluation des marges de résection saines, qui doit également être optimisée car le cancer n'est pas complètement enlevé dans un quart des cas. Compte tenu du nombre de patients atteints de carcinomes cutanés, cela augmente considérablement la charge médico-économique totale [8].

L'utilisation de méthodes optiques (également appelées biopsie optique) pour l'examen de la peau peut fournir des informations diagnostiques en quelques minutes (contre quelques jours pour l'histopathologie) et peut être utilisée *in vivo* [6, 9, 10] (évitant ainsi les procédures invasives). Ces avantages font des méthodes de biopsie optique un outil d'aide au diagnostic en temps réel prometteur pour la pratique clinique. L'intérêt des méthodes de biopsie optique réside dans leur sensibilité aux modifications des propriétés optiques de la peau à l'échelle tissulaire, cellulaire et subcellulaire, qui sont directement liées aux modifications morphologiques et métaboliques pathologiques au cours du développement des lésions cancéreuses.

Depuis plus de vingt ans, différentes méthodes de spectro-imagerie optique ont été étudiées pour la détection et/ou le diagnostic *in vivo* du cancer de la peau humaine [9, 11, 12]: (i) la spectroscopie de réflectance diffuse (DRS) [13–15], sensible à la présence d'éléments

absorbants (tels que les chromophores, la mélanine et l'hémoglobine du sang) dans la peau et aux changements morphologiques dus aux lésions cancéreuses, qui ont un impact sur la diffusion de la lumière à travers la peau (disparition de la structure en couches, désorientation des cellules, modification de la taille des noyaux des cellule), (ii) la spectroscopie d'autofluorescence (AFS) [16–18] et la microscopie en temps de vie de fluorescence (FLIM) [19] sensibles aux microenvironnements et au métabolisme cellulaire modifiés au cours de la progression tumorale, notamment grâce aux fluorophores endogènes présents dans le derme (élastine, collagène) et dans l'épiderme (flavines, nicotinamide adénine dinucléotide), (iii) la spectroscopie Raman (RS) [20–24] qui permet d'extraire des informations chimiques et cristallines, structurelles et métaboliques, liées à l'état vibratoire (diffusion inélastique) des molécules constituant les protéines et les lipides de la peau, (iv) l'imagerie hyper-spectrale (HSI) [25–27] qui combine de hautes résolutions spatiales et spectrales, fournissant ainsi des images de la peau avec des qualités spectrales améliorées pour analyser les caractéristiques d'absorption et de diffusion, (v) la microscopie confocale par réflectance (RCM) [28] et la tomographie par cohérence optique (OCT) [29–31], qui est une modalité d'imagerie en profondeur sensible aux variations de l'indice de réfraction entre les différents constituants de la peau (organites, cellules, fibres, matrice extracellulaire). L'OCT fournit des images transversales orientées verticalement au niveau cellulaire (pour l'OCT à champ linéaire) et jusqu'au derme superficiel. Cependant, il n'existe pas de méthode unique qui fournisse des informations sur les changements de toutes les propriétés optiques de la peau. Ce problème peut être partiellement résolu par l'utilisation de méthodes multimodales, qui impliquent l'utilisation simultanée de plusieurs méthodes de mesure dans une seule étude [32, 33]. Par exemple, parmi les approches multimodales les plus étudiées et testées en pratique clinique pour le diagnostic *in vivo* du cancer de la peau, la DRS et l'AFS avec une ou plusieurs excitations ont été combinés [34–36] et la RS a démontré sa contribution supplémentaire aux mesures colocalisées en DRS et l'AFS [33, 37, 38] ou dans l'OCT [39].

Notre groupe de chercheurs au sein du CRAN a développé et breveté un dispositif médical de spectroscopie multimodale [40, 41] combinant les spectroscopies DR résolues dans l'espace et AF multi-excitées, résolues spatialement qui a d'abord été mis en œuvre dans des études précliniques sur des modèles de cancer de la peau chez la souris *in vivo* [42, 43]. Récemment, notre groupe a également publié une base de données unique et entièrement ouverte de plus de 26 000 spectres DR et AF, acquis sur 131 patients en clinique, avec des lésions de BCC, SCC et de kératose actinique (AK) [44]. Par conséquent, le traitement et

l'analyse d'ensembles de données aussi complexes et multidimensionnels posent des défis majeurs. Avec le nombre beaucoup plus important de variables (c'est-à-dire de nombreuses caractéristiques par rapport au nombre d'individus et les niveaux élevés de variabilité intra- et inter-patients), des méthodes avancées d'apprentissage automatique et d'analyses statistiques sont nécessaires pour lutter contre le "fléau de la dimension", atténuer l'ajustement excessif et extraire des caractéristiques discriminantes pertinentes. Une solution à ces défis est nécessaire pour parvenir à une différenciation efficace entre les différentes classes de pathologies cutanées.

Diverses approches de classification supervisée telles que les machines à vecteurs de support [45–47], la régression logistique [14], l'analyse discriminante [34, 43] et différentes variantes des réseaux neuronaux artificiels [13, 17] ont été mises en œuvre pour identifier les différents types de cancers de la peau sur la base des spectres AF et DR. Cependant, les performances des classificateurs peuvent varier de manière significative en fonction du réglage des paramètres et des données à analyser ; il est donc nécessaire de comparer les performances de plusieurs classificateurs et d'effectuer une optimisation des hyperparamètres pour chacun d'entre eux afin d'améliorer les résultats.

Un autre aspect important est l'utilisation efficace des données multimodales. À cette fin, il est possible d'exploiter les méthodes de fusion de données, qui ont été testées avec succès dans des études précliniques [46, 47].

La combinaison adéquate des méthodes spectroscopiques d'examen de la peau avec les méthodes de traitement, de classification et de fusion des données est une tâche assez complexe, qui nécessite un grand nombre de tests préliminaires et, par conséquent, des coûts importants en termes de temps. Ceci est particulièrement pertinent dans le contexte des essais cliniques sur des sujets humains, étant donné que la préparation de tous les documents et expériences nécessaires prend souvent jusqu'à plusieurs dizaines d'années. À cet égard, le développement de méthodes d'analyse des données obtenues sur les données des études précliniques est généralement la meilleure solution.

En tenant compte de toutes les difficultés mentionnées, les objectifs suivants ont été définis pour cette thèse :

1. Développer une méthodologie pour le traitement et la classification des données spectroscopiques multimodales, y compris l'utilisation de techniques d'apprentissage automatique, sur les spectres obtenus dans une étude préclinique sur la peau de souris avec des lésions bénignes et potentiellement précancéreuses.

2. Évaluer l'efficacité de la méthodologie de traitement des données spectroscopiques multimodales développée sur des données obtenues dans le cadre d'une étude clinique sur la peau de personnes atteintes d'un cancer de la peau sans mélanome ou de lésions précancéreuses.
3. Déterminer les moyens possibles d'améliorer le traitement des spectres et la méthodologie de classification qui peuvent être utilisés dans les étapes suivantes des études cliniques pertinentes.

Contexte spécifique de la thèse

Cette thèse a été financée par l'ambassade de France en Russie dans le cadre d'une bourse Vernadski de cotutelle internationale avec l'Université de Tomsk (2021-2024) et par l'Université de Lorraine (Lorraine Université d'Excellence, projet PIA financé par l'ANR, ANR-15-IDEX-04-LUE). Les données utilisées dans cette thèse ont été obtenues dans le cadre du projet Spec-LCOCT financé par l'ANR (ANR-21-CE19-0056) et ont été réalisées avec le dispositif SpectroLive de la plateforme PhotoVivo, qui fait partie du réseau France Life Imaging (FLI) et financé par le Contrat de Plan Etat-Région Grand Est 2015-2020 (CPER IT2MP : Innovations Technologiques, Modélisation et Médecine Personnalisée) remercie le soutien financier du Fonds Européen de Développement Régional (FEDER), de la Région Grand Est et de la Ligue Contre le Cancer.

Développement du manuscrit

La thèse est organisée comme suit :

Le chapitre 1 passe en revue les propriétés et la structure des différentes couches de la peau, ainsi que les informations de base sur les carcinomes cutanés et les cancers de la peau. Il décrit les principales étapes de la procédure de diagnostic du cancer de la peau utilisée en clinique et examine l'utilisation des méthodes optiques pour le diagnostic du différentiel des carcinomes cutanés et des kératoses actiniques, y compris les méthodes multimodales.

Le chapitre 2 passe en revue les principales étapes du prétraitement et du pipeline de traitement qui peuvent être utilisées pour les données spectroscopiques, ainsi que les méthodes les plus importantes qui peuvent être utilisées pour chaque étape, y compris les méthodes d'apprentissage automatique et de fusion de données. En outre, après avoir décrit le traitement des données, ce chapitre aborde également les principales mesures de précision qui peuvent être utilisées pour évaluer les performances de l'analyse.

Le chapitre 3 décrit le matériel et les méthodes de l'étude préclinique sur la peau de souris ainsi que les résultats obtenus. Une attention particulière est accordée à la comparaison des performances de différentes combinaisons de techniques de réduction de la dimensionnalité, de classification et de fusion de données.

Le chapitre 4 décrit le matériel et les méthodes de l'étude clinique sur la peau humaine ainsi que les résultats obtenus. Une attention particulière est accordée à la comparaison des performances des différentes modalités ainsi qu'à la comparaison des résultats obtenus à la fois avec les résultats de l'étude préclinique décrite au Chapitre 3 et avec les résultats obtenus par les études diagnostiques des carcinomes utilisant des méthodes optiques unimodales et multimodales décrites au Chapitre 1.

La conclusion générale et les perspectives résument les principaux résultats de l'étude présentée et discutent des étapes possibles pour son amélioration future.

General Introduction

Context of the work

In 2022, non-melanoma skin cancer (NMSC) was the fifth most common cancer type in the world [1]. Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC) are the most common types of NMSCs, accounting for 70% and 25% of cases respectively. Metastasis occur in only 1 out of 14,000,000 cases for BCC and in less than 10 % for SCC [2], inducing an overall 90%-five-year survival rate. The age-standardized incidence rate of NMSC increased from 54.08/100,000 in 1990 to 79.10/100,000 in 2019 (+1.78 % per year) and, the number of reported cases is increasing every year [3–5]. This increasing incidence entails an important medico-economic burden.

Indeed, the standard procedure for of suspicious skin lesion diagnosis includes surgical biopsy and histopathological analysis of the removed tissue samples [6, 7]. Thus, according to [8], the ratio of biopsies to treatment increased from 1,1 in 1993 to 2,1 in 2016. In addition, biopsy is invasive, long and expensive procedure which induce stress and scars and has low diagnostic accuracy, since 20% of the biopsies only reveal skin cancers and two-thirds of the carcinomas are not diagnosed.

Another important aspect is the assessment of healthy resection margins, which must also be optimized as the cancer is not completely removed in one quarter of the cases. Considering the number of patients with NMSC, this significantly increases the total economic burden [8].

The use of optical methods (also refer to as optical biopsy) for skin examination may provide diagnostic information within a few minutes (versus a few days for histopathology) and can be used *in vivo* [6, 9, 10] (thus avoiding invasive procedures). These advantages make the optical biopsy methods to be a promising diagnostic real time assistance tool for clinical practice. The interest of optical biopsy methods lies in their sensitivity to modifications in the optical properties of the skin at the tissue, cellular and subcellular scales which are directly linked to pathological morphological and metabolic modifications during cancerous lesion development.

For more than two decades now, different optical spectro-imaging methods have been investigated for human skin cancer *in vivo* detection and/or diagnosis [9, 11, 12], among which: (i) diffuse reflectance spectroscopy (DRS) [13–15] sensitive to the presence of absorbing elements (such as chromophores, melanin and blood haemoglobin) in the skin and to

morphological changes due to cancerous lesions, which impact the elastic scattering of light (disappearance of the layered structure, disorientation of cells, modification of size nuclei), (ii) autofluorescence spectroscopy (AFS) [16–18] and fluorescence lifetime imaging microscopy (FLIM) [19] sensitive to microenvironments and cellular metabolism altered during tumor progression, namely through endogenous fluorophores present in dermis (elastin, collagen) and in epidermis (Flavins, Nicotinamide adenine dinucleotide), (iii) Raman spectroscopy (RS) [20–24] which makes it possible to extract chemical and crystalline, structural and metabolic information, which is linked to the vibrational state (inelastic diffusion) of the molecules constituting the proteins and lipids of the skin, (iv) Hyper-Spectral Imaging (HSI) [25–27] which combines high spatial and spectral resolutions, thus providing skin pictures with enhanced spectral qualities for analyzing absorption and scattering features, (v) Reflectance Confocal Microscopy (RCM) [28] and Optical Coherence Tomography (OCT) [29–31], which is an in-depth imaging modality being sensitive to the refractive index variations between the various constituents in skin (organelles, cells, fibers, extracellular matrix). OCT provides vertically oriented cross-sectional images at cellular level (for Line-field OCT) and down to superficial dermis. However, there is no single method that provides information about changes in all skin optical properties. This issue can be partially solved by using multimodal methods, which involve the simultaneous use of several measurement methods in a single study [32, 33]. For instance, among the most studied multimodal approaches tested in clinical practice for *in vivo* skin cancer diagnosis, DRS and AFS with one and several excitations were combined [34–36] and RS demonstrated its further contribution to colocalized measurements in DRS and AFS [33, 37, 38] or in OCT [39].

Our group of the CRAN laboratory has developed and patented a multimodal spectroscopy medical device [40, 41] combining spatially-resolved DR and multiply-excited AF spectroscopies which was implemented first in preclinical studies on mice skin cancer models *in vivo* [42, 43]. Recently, our group also published a unique and fully open database of more than 26000 DR and AF spectra, acquired on 131 patients in clinics, with BCC, SCC and Actinic Keratosis (AK) lesions [44]. Consequently, processing and analyzing such complex and multi-dimensional datasets involve major challenges. With the much larger number of variables (i.e., numerous features compared to the number of individuals and the high levels of intra- and inter-patient variabilities) advanced methods of Machine Learning (ML) and statistics are necessary to tackle the curse of dimensionality, to mitigate overfitting and to extract relevant

discriminant features. A solution for these challenges is required to reach an efficient differentiation between different classes of skin pathologies.

Various supervised classification approaches such as Support Vector Machine (SVM) [45–47], Logistic Regression [14], Discriminant Analysis [34, 43] and different variations of Artificial Neural Networks (ANN) [13, 17] have been implemented to identify different types of skin cancers based on AF and DR spectra. However, the performance of classifiers can significantly vary depending on the parameter adjustments and the data to be analyzed; therefore, it is necessary to compare the performance of several classifiers and perform hyperparameter optimization for each of them in order to improve results.

Another important aspect is the effective use of multimodal data. For this purpose, it is possible to exploit Data Fusion methods, which have been successfully tested in preclinical studies [46, 47].

The proper combination of spectroscopic methods of skin examination with data processing, classification and data fusion methods is a rather complex task, requiring a large number of preliminary tests and, consequently, large time costs. This is especially relevant in context of clinical trials on human subjects, as the preparation of all necessary documents and experiments often takes up to decades. In this regard, the development of methods for analysing the obtained data on the data of preclinical studies is usually the best solution.

Taking into account all the mentioned difficulties, the following objectives have been defined for this thesis:

1. To develop a methodology for processing and classification of multimodal spectroscopic data, including the use of machine learning techniques, on spectra obtained in a preclinical study on the skin of mice with benign and potentially precancerous lesions.
2. To evaluate the effectiveness of the developed multimodal spectroscopic data processing methodology on data obtained in a clinical study from the skin of people with non-melanoma skin cancer or precancerous lesions.
3. To determine possible ways to improve the spectra processing and classification methodology that can be used in the following stages of the relevant clinical study.

Specific context of the work

This thesis was funded by the French embassy in the Russian Federation under a Vernadski international joint PhD grant (2021-2024) completed by the Université de Lorraine

and the French National Research Agency (ANR) under the French PIA project “Lorraine Université d’Excellence” (ANR-15-IDEX-04-LUE). The data used in this thesis were obtained as part of the Spec-LCOCT project funded by the ANR (ANR-21-CE19-0056) and was carried with the SpectroLive device of the PhotoVivo platform, which is part of the France Life Imaging (FLI) network and funded by Contrat de Plan Etat-Région Grand Est 2015-2020 (CPER IT2MP: Innovations Technologiques, Modélisation et Médecine Personnalisée) thank s to financial support by European Regional Development Fund (FEDER), Grand Est Region and Ligue Contre le Cancer.

Thesis outline

The thesis is organized as follows:

Chater 1 review the properties and structure of different skin layers, basic information on skin carcinomas and AK. It describes the main steps of the skin cancer diagnostic procedure used in the clinic, and reviews the use of optical methods for NMSC diagnosis, including multimodal methods.

Chapter 2 reviews the main steps of pre-processing and processing pipeline that can be used for spectroscopic data, as well as the most important methods that can be used for each step, including machine learning and data fusion methods. In addition, after providing the description of processing of data, this chapter also discusses the main accuracy metrics that can be used to assess the performance of its analysis.

Chapter 3 describes the materials and methods of the preclinical study on mice skin as well as the results obtained. Particular attention is paid to comparing the performance of different combinations of dimensionality reduction, classification and data fusion techniques.

Chapter 4 describes the materials and methods of the clinical study on human skin as well as the results obtained. Particular attention is paid to comparing the performance of different modalities as well as comparison of the results obtained both with the results of the preclinical study described in Chapter 3 and with the results obtained by NMSC diagnostic studies using unimodal and multimodal optical methods described in Chapter 1.

General conclusion and Prospects summarizes the main results of the presented study and discusses possible steps for its future improvement..

The list of scientific contributions

International journals

1. V. Kupriyanov, W. Blondel, C. Daul, M. Amouroux, and Y. Kistenev, “Implementation of data fusion to increase the efficiency of classification of precancerous skin states using in vivo bimodal spectroscopic technique,” J. Biophotonics, vol. 16, no. 7, p. e202300035, Jul. 2023, doi: 10.1002/jbio.202300035 hal-04089067

2. Elsen T., Fauvel C., Khairallah G., Zghal A., Delconte A., Kupriyanov V., Blondel W. and Amouroux M. (2024) “A dataset of optical spectra and clinical features acquired on human healthy skin and on skin carcinoma,” *Data in Brief*, 53:110163. https://doi.org/10.1016/j.dib.2024.110163 hal-04458411

Under submission

3. V. Kupriyanov, W. Blondel, C. Daul, G. Khairallah,, Y. Kistenev and M. Amouroux, “Machine learning based-classification of spatially resolved diffuse reflectance and autofluorescence spectra acquired on human skin for actinic keratoses and skin carcinomas diagnostics aid,” submitted to Journal of biomedical optics.

International conferences

1. V. Kupriyanov, W. Blondel, C. Daul, M. Amouroux, Y. Kistenev, and D. Vrazhnov (2022) “Diagnosis of skin precancerous stages in an in vivo preclinical model using bimodal spectroscopy and machine learning,” in 16th International conference on Laser Application in Life Science, LALS 2022, Nancy, France, Apr. 2022. Accessed: Jul. 01, 2024. [Online]. Available: https://hal.science/hal-03893848 hal-03893848 *(Poster)*

2. V. Kupriyanov, W. Blondel, C. Daul, M. Amouroux, Y. Kistenev (2022) “Diagnosis of benign and precancerous skin conditions using bimodal spectroscopy, machine learning and Data Fusion methods” in 10th International Symposium "Optics and Biophotonics", SFM'22, Saratov, Russia, September 26 – 30. Available: https:// https://hal.science/hal-03893848v1 *(On-line Report)*

3. V. Kupriyanov, W. Blondel, C. Daul, M. Amouroux, and Y. Kistenev (2023) “Assessing the effectiveness of multimodal data fusion methods for in-vivo diagnosis of pre-cancerous skin conditions in a preclinical model,” (2023) In *SPIE Proceedings, Optical Biopsy XXI*, 12373-24, SPIE Photonics West, January 28-February 2, San Francisco, USA, pp. 34–38.

doi: 10.1117/12.2649626 hal-03979475 *(Oral Report)*

4. V. Kupriyanov, W. Blondel, C. Daul, M. Amouroux, and Y. Kistenev (2023) "Evaluating the efficacy of different data processing methods in diagnosing precancerous skin conditions in a preclinical in-vivo model using bimodal spectroscopy," In *SPIE Proceedings, Dynamics and Fluctuations in Biomedical Photonics XX*, 12378-26, SPIE Photonics West, January 28-February 2, San Francisco, USA, pp. 49–52. doi: 10.1117/12.2650369 hal-03979509 *(Poster)*

5. A. Zghal, Fauvel C., Kupriyanov V., Elsen T., Khairallah G., Blondel W. and Amouroux M. (2023) "Human skin phototype and apparent age classification based on machine learning methods of autofluorescence and diffuse reflectance spectroscopic data acquired in vivo," In *SPIE Proceedings, Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XXI*, 12368-32, SPIE Photonics West, January 28-February 2, San Francisco, USA, pp. 60–64. doi: 10.1117/12.2649661 hal-03979527 *(Oral Report)*

6. V. Kupriyanov , M. Pinheiro, S. Carvalho, I. Carneiro, R. Henrique, V. Tuchin, L. Olivera, M. Amouroux, Y. Kistenev and W. Blondel "The use of data mixing to increase the size of the dataset for colon cancer diagnosis using diffuse reflectance spectroscopy and machine learning" in 11th International Symposium "Optics and Biophotonics", SFM'23, Saratov, Russia, September 25 – 29. hal-04708401

7. Blondel W., Kupriyanov V, Daul C., Khairallah G., Kistenev Y. and Amourourx M. (2023) "Multimodal optical spectroscopy for human skin cancer in vivo diagnosis" In 17th International conference on Laser Applications in Life Sciences LALS2023, Muğla, Turkey, Oct. 15-17. hal-04246046 (*Invited Lecture)*

8. Blondel W., Kupriyanov V., Daul C., Khairallah G., Kistenev Y. and Amouroux M. (2023) "Multimodal optical spectroscopy for human skin cancer in vivo diagnosis", In *Asia Communications and Photonics Conference - The International Photoniocs and Optoelectronics Meetings ACP-POEM 2023, 10th Sino-French PHOTONET International Research Network Workshop*, Wuhan, China, Nov. 04-07. hal-04271184 (*Invited Lecture)*

9. Kupriyanov V., Blondel W., Daul C., Khairallah G., Fauvel C., Kistenev Y. and Amouroux M. (2024) "Bimodal spectroscopy for skin carcinomas and actinic keratoses diagnostic assistance", In *SPIE Proceedings, Clinical Biophotonics III*, 13009-3, SPIE Photonics Europe, April 7-11, Strasbourg, France. https://doi.org/10.1117/12.3016431, hal-04634323. *(Oral Report)*

10. Kupriyanov V., Amouroux M., Daul C., Khairallah G., Fauvel C., Kistenev Y. and Blondel W. (2024) “Estimation of skin age and phototype using bimodal spectroscopy and machine learning”, In *SPIE Proceedings, Tissue Optics and Photonics III*, 13010-28, SPIE Photonics Europe, April 7-11, Strasbourg, France. https://doi.org/10.1117/12.3016417, hal-04634341 *(Poster)*

11. Kupriyanov V., M. Pinheiro M., Carvalho S., Carneiro I., Henrique R., Tuchin V., Olivera L., Amouroux M., Kistenev Y. and Blondel W. (2024) “Classification of healthy and cancerous colon tissues based on absorption coefficient spectra”, In *SPIE Proceedings, Tissue Optics and Photonics III*, 13010-53, SPIE Photonics Europe, April 7-11, Strasbourg, France., https://doi.org/10.1117/12.3021894, hal-04634348. *(Oral Report)*

National conferences / workshops

1. Kupriyanov V., Blondel W., Daul C., Amouroux M. and Kistenev Y. (2022) “Comparison of the performance of different classification and decision fusion methods to analyse spectra obtained from precancerous skin lesions by using bimodal spatially resolved spectroscopic technique.” In *Journée scientifique du GDR ISIS Apprentissage automatique multimodal et fusion d'informations (3ième édition)*, Paris, France, Dec 15, 2022. hal-03893899 *(Oral Report)*

.

Chapter 1: Review of existing methods for the diagnosis of non-melanoma skin cancer

1.1. Introduction

Development of new methods for non-invasive diagnosis of non-melanoma skin cancer and precancerous lesions is an important task of modern science, as according to the number of diagnosed cases is increasing every year. The main disadvantages of traditional NMSC diagnosis methods are usually associated with the need for surgical biopsy, which is an invasive procedure and requires large time and economic costs.

Optical methods can be a good alternative to conventional skin biopsy as they can be applied non-invasively and are highly sensitive to changes in skin properties and structure due to carcinogenesis.

This chapter will review the properties and structure of different skin layers, basic information on skin carcinomas and AK, a description of the main steps of the skin cancer diagnostic procedure used in the clinic, and an overview on the use of optical methods for NMSC diagnosis, including multimodal experimental methods.

1.2. Clinical methods of skin cancer diagnosis

1.2.1. Skin structure and composition

The human skin is the largest organ of the body and accounts for about 10 per cent of its total mass. Its main function is to protect the human body from external damages caused by ultraviolet radiation, chemicals, allergens or microorganisms. The skin is important in regulating body temperature and blood pressure and in addition, it is a sensory organ [48].

Human skin is a complex heterogeneous structure, which consists of three main layers: epidermis, dermis and adipose tissue [48, 49]. The general scheme of skin layers is shown in **Figure 1.1**. The composition and main functions of each layer are presented in more detail in the next sections.

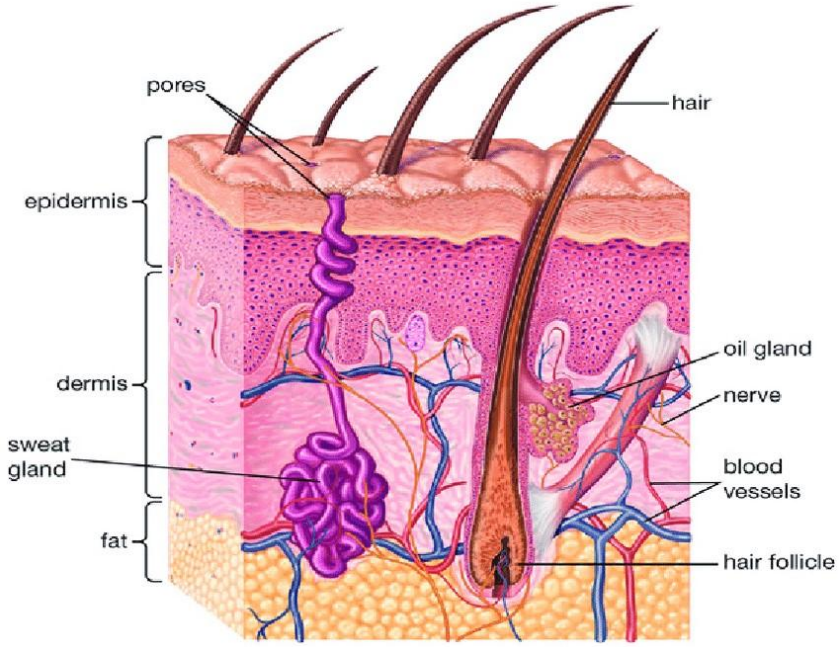


Figure 1.1. Scheme of the human skin layers [49, 50].

1.2.1.1. Epidermis

The epidermis is the outer layer of the skin, with a thickness of ten to hundreds of micrometres, and acts as a barrier [51, 52]. The epidermis can also be divided into several sub-layers, which are shown in the **Figure 1.2** [52].

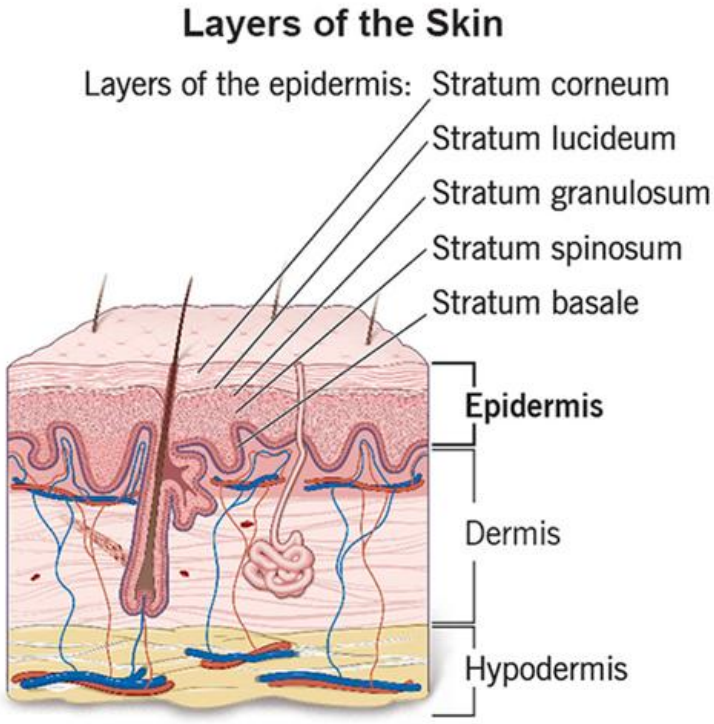


Figure 1.2. Scheme of the epidermis sub-layers in human skin [52].

The main outermost layer of the epidermis responsible for barrier function is the **Stratum Corneum (SC)**. It is mostly composed of keratin (a fluorescent protein) and corneocytes, which are a dead keratinocytes. The thickness of the stratum corneum is about 20 micrometres in average, which is equivalent to about 20 up to 30 layers of flattened dead cells. However, the thickness of the stratum corneum can vary from 10 to 40 μm depending on the mechanical stress on the concerned body area [51, 53].

The next layer of the epidermis is called Stratum lucidum. It is much thinner than the SC later and consists of only about 2-3 cell layers. This layer is only prominent on the palms of the hands and soles of the feet and provides additional protection against friction, which is necessary for these areas of the skin [53].

Stratum granulosum or granular layer consists of 3-5 layers of keratinocytes with keratohyalin granules and lamellar bodies. Keratohyalin granules are cytoplasmic structural proteins which contain keratin precursor that eventually aggregate, cross-link and form bundles [52, 53]. The lamellar bodies contain glycolipids that are secreted onto the cell surface and keep the cells connected to each other. On the border with the Stratum Corneum, lamellar bodies are secreted from keratinocytes into the intercellular space, while losing their nuclei and organelles. Subsequently, this leads to the corresponding formation of dead keratin cells and lipid bilayer in the SC.

Between the basal layer and the granular layer is the Spiny layer, which is 8-10 cell layers thick. It is in this layer of the epidermis that keratin is synthesised. Immunologically active Langerhans cells are also located in this layer [54].

The bottom layer of the epidermis is the basal layer, separated from the dermis by the basal membrane and attached to it by junction structures called hemidesmosomes. The cells found in this layer are mitotically active stem cells of cubic to columnar shape that continuously produce keratinocytes. This layer also contains melanocytes that produce melanin (one of the main chromophore in human skin) [53].

1.2.1.2. Dermis and Hypodermis

The next skin layer is the dermis, which represents the major part of the skin volume and is mainly responsible for its trophic and supportive functions [53, 55]. The dermis is from 2 to 5 mm thick and consists of two main layers, the papillary dermis and the reticular dermis, in which blood and lymphatic vessels can be found. It's scheme is presented in **Figure 1.3** [56].

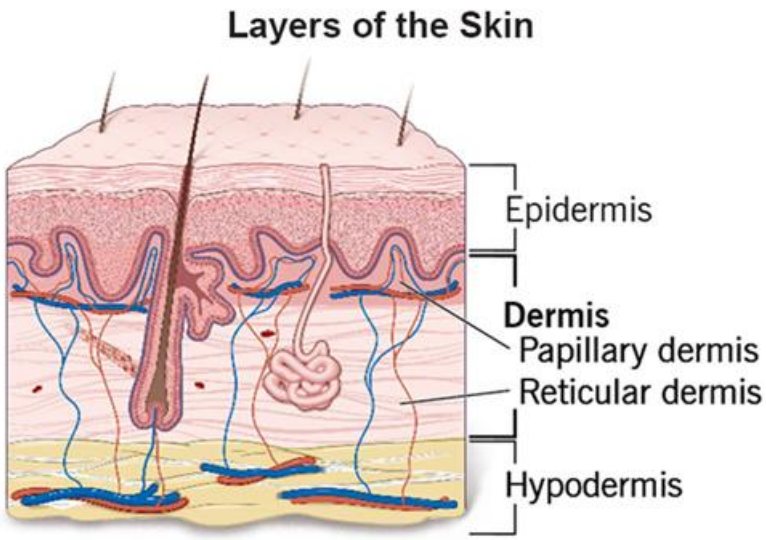


Figure 1.3. Scheme of the dermis sub-layers and hypodermis structure in human skin [56].

The papillary dermis, which gets its name from the "papillae" - the outgrowths by which this layer of the dermis connects to the basal membrane, consists mainly of collagen fibres (one of the two main fluorescing structural protein in skin) and blood vessels - capillaries, that help to supply the epidermis with nutrients and regulate skin temperature. The network of elastin fibres (the other main fluorescing structural proteins in skin) is also mainly located in this layer. The main function of the papillary dermis is to make the skin elastic, stretchable and more stable [53, 55, 56].

The reticular layer is the bottom layer of the dermis, which extends from the papillary dermis to the hypodermis. This layer consists of collagen fibres that form a mesh-like structure which provides additional elasticity to the skin [55, 56]. In addition, the reticular dermis contains blood vessels, hair follicles, lymphatic vessels nerves and fat cells [55, 57].

The hypodermis or subcutaneous tissue is an inner layer that consists of fat, connective tissue, larger calibre blood vessels, nerve endings, and lymphatic vessels. The main functions of the hypodermis include the accumulation of fat cells and protecting the body from cold, heat, as well as bumps and falls. In addition, blood vessels and nerves are connected to the skin through the hypodermis [58].

1.2.2. Skin cancers

According to the International Agency for Research on Cancer [59], the total number of reported cases of skin cancer in 2022 was over 1,500,000, representing 7.8% of the global

number of reported cases of all types of cancer. Skin cancer can be divided into two main types: Melanoma Skin Cancer (MSC) and Non-Melanoma Skin Cancer (NMSC).

MSC accounts for 21% of all reported skin cancer cases worldwide. At the same time, MSC is responsible for 58667 deaths in 2022, which represents 46% of all skin cancer-related deaths. Melanoma develops from melanocytes and is most often associated with sun exposure, mainly intermittent and intense during childhood. In addition, pigmented nevi may be precursors to melanoma.

NMSC is the fifth most common type of cancer in the world and occurs several times more frequently than melanoma. At the same time, the number of deaths caused by this type of skin cancer, normalised by the total number of cases, is more than 3 times lower than for melanoma (5.6% vs. 17.9%, respectively). NMSC develops from keratinocytes, and has two main forms: Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC), which will be described in more detail in the following sections. The main cause of NMSC, as for MSC, is sun exposure. In addition, SCC can also develop from Actinic Keratoses.

In addition to the immediate increase in tumour size, skin cancer can also be invasive, meaning that it can metastasise to surrounding tissues and organs. In such cases, more serious and urgent treatment may be required, and the patient's likelihood of survival is reduced. All this makes it necessary to accurately determine the invasiveness of the lesion. The risk of invasiveness for melanoma (15%) is higher than for NMSC (from 2 to 14%).

1.2.2.1. Basal Cell Carcinoma

Basal Cell Carcinoma (BCC) accounts for at least 32% of all cases of cancer worldwide and about 70% of all NMSC cases, making it the most common cancer worldwide. According to most scientists [60–62], the most significant factor in the occurrence of BCC is chronic and intensive UltraViolet (UV) light exposure. However, in contrast to melanoma, the chronic nature of the exposure is more important than its intensity. At the same time, although exposure to ultraviolet radiation is accepted as a critical causative factor in the pathogenesis of basal cell carcinoma, the magnitude of the risk associated with increased exposure seems to be insufficient to explain either why particular people get these tumours whereas others do not, or the considerable phenotypic diversity shown by patients in terms of the number and site of tumour and patterns of presentation. Other risk factors include heredity, radiation exposure, burns, scarring, chronic inflammation and others. The likelihood of getting BCC is lower for people with darker skin colour. The average age of patients suffering from BCC is over 60 years

[4, 63]. BCC is characterised by slow growth and often develop over many months or even years.

BCC can be located on different anatomical regions of the body, but about 80% of all reported cases occur on the head and neck (as a consequence of higher sun exposure), with the remaining BCC mainly lying on the trunk and lower limbs [64]. At the same time, the risk of finding BCC on the backs of the hands is much lower. At the initial stage of development, BCC lesions are commonly small, translucent or pearly, with raised areas through which dilated vessels may show [65]. The classic form is the rodent ulcer, which has an indurated edge and ulcerated centre [66]. The lesion grows slowly, but if it is neglected, it can spread deep and cause severe damage, especially if it is located in the eye, nose or ear area. The periorbital tissues and bones may also be affected.

There are different subtypes of BCC including nodular or cystic BCC, superficial BCC, morphoeic BCC and pigmented BCC [67]. Superficial BCC tend to occur on the trunk. They are often flat, well demarcated erythematous plaques, which can mimic psoriasis, discoid eczema, and Bowen's disease. Nodulocystic basal cell carcinoma presents as a solitary, shiny, red nodule with large telangiectatic vessels, often seen on the face.

Patients with BCC lesions have an increased risk of developing other BCCs in future. In a recent meta-analysis, the three-year cumulative risk of new cases of BCC varied between 33% and 77%, which may directly correlate with the number of lesions observed. In addition, patients with BCC detected in the trunk area have a higher risk of having recurrent lesions (hazard ratio 1.58) [64, 68, 69].

1.2.2.2. Squamous-Cell Carcinoma

SCC (Squamous-Cell Carcinoma) or cSCC (cutaneous Squamous-Cell Carcinoma) usually appears as a hard growth with a scaly top, but can also form into an ulcer [70]. The development of the disease often occurs within a few months. As for BCC, the most influential cause of SCC is exposure to UV radiations [71]. Ionizing radiation has also been implicated in the pathogenesis of SCC as it was used to treat many skin diseases, such as acne, dermatitis, and hemangiomas. Another important risk factor for SCC is the influence of chemical agents [21].

SCC *in situ* may progress to invasive disease when it's not completely treated in time. Invasive SCC is characterized by invasion of the dermis by the neoplastic squamous epithelial cells [71]. The invasiveness can take the form of infiltrating cords, sheets, or single cells, or can

present as well-circumscribed nodules, squamous islands, or cystic structures composed of malignant keratinocytes. Invasive SCC is more difficult to diagnose and treat.

Non-invasive forms of SCC include Erythroplasia of Queyrat and Bowen's disease [71, 72]. Bowen's disease is a form of cancer that develops only on the surface of the skin. It usually occurs in old women and is characterised by the appearance of one or more distinct and irregularly shaped red plaques on the trunk or legs. Erythroplasia of Queyrat is histologically identical to Bowen's disease, but the neoplasm appears on the head of the penis, the labia minora or the inner foreskin [71]. It is usually a red spot or plaque with clear borders and a shiny moist surface. During carcinogenesis, the surface of the lesion becomes ulcerated and bleeds profusely even with minor mechanical injuries.

For invasive SCC, a distinction is made between exophytic (tumour-like) and endophytic (ulcerated) forms of cancer [73]. The exophytic form of growth is covered by a crust or additional layers of SC cells and is usually represented by a plaque or nodule adherent to the underlying tissues measuring more than 2 cm. As the tumour progresses, ulcers appear on the surface of the tumour, indicating that it has become endophytic.

Unlike BCC, SCC has a higher risk of metastasis and may spread to local lymph nodes [74]. The prognosis for SCC depends on several factors: the carcinoma subtype, the treatment methods used, the location and severity of the tumour, and various variables related to the patient’s condition (comorbidities, age, etc.). The long-term prognosis is generally positive, as less than 4 % of cases of squamous cell carcinoma are at risk of metastasis [74]. If metastasis does occur, regional lymph nodes are involved in approximately 85 percent of all cases; approximately 15 percent of cases involve distant sites, including the lungs, liver, brain, skin, and bone. Ten-year survival rates are less than 20 percent for patients with local lymph node involvement and less than 10 percent for patients with distant metastases.

1.2.2.3. Actinic Keratoses

Actinic Keratoses (AK) are precancerous neoplasms that, like skin carcinomas, occur predominantly on sun-exposed areas of the skin. Actinic keratoses can appear either as papules whose size vary from 1 to 2 mm, or as large plaques [75]. In most cases, the diagnosis of AK is based on the clinical appearance of the lesion. However, it can sometimes be almost impossible to differentiate it from SCC and other pathological lesions without biopsy and subsequent histological examination.

Depending on the degree of development there are several classifications of AK, however, almost all of them contain three stages. Olsen *et al.* recommended assessing the severity of lesions by clinical presentation [76]. Another classification proposed by the European Guidelines for the Treatment of AK and created by Röwert-Huber *et al.* arranges the lesions according to the results of histological examination [77]. It is important to emphasise that the degree of correspondence between clinical and histological grading is low, which confirms the need to treat each AK regardless of the grade.

Actinic Keratosis is also a possible precursor to SCC [78, 79]. They can involute or persist, and affected persons usually have many lesions, some of which may evolve into SCC. The cumulative risk of having at least one invasive SCC in future varies from 2 to 6% in patients with one lesion to 20,9% in patients with severe AK lesions [80]. Actinic keratosis has been described as a type of carcinoma *in situ*, more precisely as a carcinoma involving only the epidermis. From a therapeutic point of view, it may be impractical and unnecessary to treat each individual keratotic lesion, but patients with many lesions should be followed closely so that evolving SCC can be detected and treated in time.

If AK development is detected, the increased risk of new cases of AK and SCC in the area close to the lesion focus, which is usually called Field Cancerization, should also be considered, as this area is exposed to the same factors (especially UV exposure) as the lesion itself.

1.2.3. Standard clinical procedure for skin cancer diagnosis and treatment

The direct determination of the type of skin lesion and further treatment is usually done by a dermatologist. However, in some cases the diagnosis may be made by a general practitioner, plastic surgeon or nurse [81, 82]. However, the first stage of the diagnostic procedure for any specialist is the visual examination of the skin (lesion) combined with the collection of anamnesis. The list of parameters, that should be evaluated and taken into account when diagnosing skin carcinoma is presented in **Table 1.1** [82].

Table 1.1. Parameters used for NMSC diagnosis [82].

Parameters obtained from the literature [83–86]	**Additional parameters mentioned by doctors**
Patient related	
Age	Previous skin cancer cases
Gender	Frequency of sun exposure
Skin type	Frequent sun tanning
Excessive sun exposure	Frequent holidays with sun exposure
Professional outdoors	Sunburned during lifetime
Immunosuppressant use	Solarium use (and frequency)
Patients after organ transplantation	Skin cancer in direct family
Lesion related	
Color	
Scaliness	
Keratotic surface	
Confinement	
Elevation	
Size	Itching
Speed of enlargement	Shininess
Induration	Presence of other signs of sun damage
Ulceration	
Pain	
Localization	
Bleeding	
Widened blood vessels	

The most important parameters in making the diagnosis are "lesion related" such as colour, size, bleeding, pain, elevation and the dynamics of the lesion. At the same time, taking into account the other parameters can also help to make a more accurate diagnosis, as the chance of different types of skin cancer correlates with age, gender, phototype, etc. However, even considering all of the listed parameters may not always be sufficient without histological analysis of the lesional tissues. In most of the clinical studies found, the values of accuracy, Sensitivity (Se) and Specificity (Sp) exceed 80% [81, 82, 87–89]. At the same time, the study [90] had a sensitivity of less than 40%, which makes it impossible to state clearly that the accuracy of clinical trials is sufficiently high in all cases.

1.2.3.1.Dermoscopy

The standard visual examination can also be supplemented by the use of dermoscopy, a non-invasive method of examining the skin using a special device called a dermoscope. The

parts of dermoscope include the lighting or illumination system that usually consists of light-emitting diode sources, magnification using achromatic lenses located in the faceplate of the dermoscope and a power supply. All dermoscopic systems can be divided into polarisation and non-polarisation types [91].

The principle of polarisation dermoscopy is to use two polarisation filters on the source and detector held orthogonally at 90°. Light reflected directly from the skin surface retains its polarisation and is completely blocked by the polarisation filter of the detector. In turn, the light reflected from deeper skin structures loses its polarisation and phase and therefore passes through the second filter. This configuration of the dermoscope allows the visualisation of deeper structures of the skin up to a 100 micron depth [91].

Nonpolarized dermoscope of any type should be used in contact with the skin surface. It provides better visualization of superficial structures such as comedones, milia-like cysts, crypts, fissures, and scales. In turn, polarized dermoscopy is more conspicuous in viewing deeper structures such as pigment network and vessels [91, 92]. Hence, both the modes are complementary to each other for the better visual perspective of the lesions.

According to [93], the use of dermoscopy for NMSC diagnosis improves diagnostic accuracy from 3 to 5% and can also narrow down the list of pathologies in differential diagnosis.

1.2.3.2. Biopsy and/or surgical treatment

In most cases, after a visual examination and history taking, a biopsy is performed to take tissue samples for further histological analysis. The biopsy can be incisional, when only a part of the lesion has to be removed to confirm the diagnosis, or excisional, when the entire lesion is completely and deeply removed till the subcutis plane. Excisional biopsy is preferred when a neoplasm is suspected [94, 95]. There are various methods of skin biopsy, and the choice of any particular technique depends on the type of lesion, the localisation of the lesion, and the qualifications of the dermatologist. However, regardless of the type of biopsy, the procedure is invasive and rather painful.

In puncture biopsy the lesion or area to be biopsied is punched out using a disposable or sterilizable punch of varying sizes [94]. For non-facial lesions, a 4-mm punch is sufficient; however, in granulomatous conditions or conditions with atypical features, biopsies of 5 mm or more are preferable. It is advisable not to take less than 3 mm biopsies as vital features may be missed in too small biopsies. The area is anaesthetized and the skin is stretched in a direction perpendicular to the resting skin lines. Advantages of punch biopsy include ease of performance

and obtaining uniformly shaped tissue. Disadvantages are that the material obtained may be inadequate and often biopsy may not include deeper tissue. An example of punch biopsy is shown in **Figure 1.4** [96].

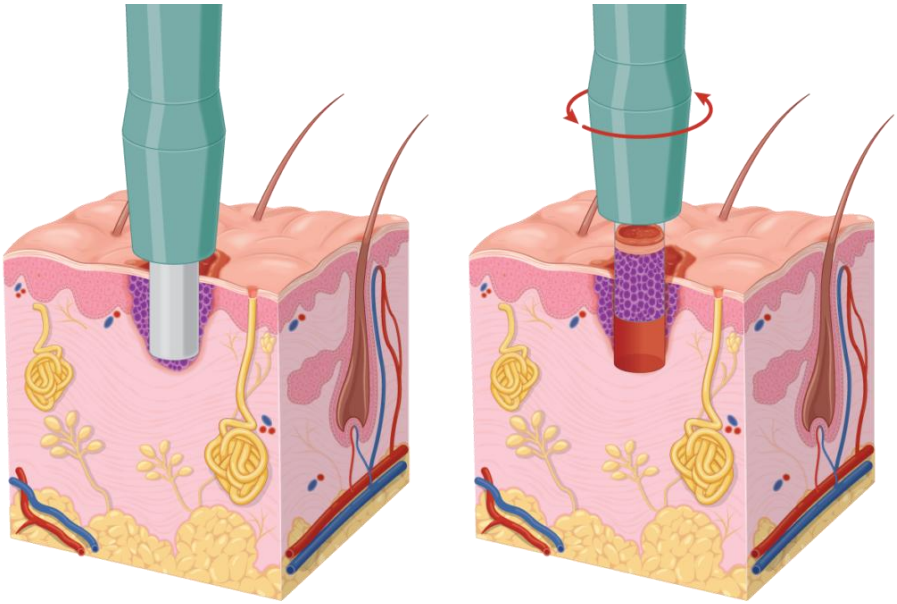


Figure 1.4. The principle of Punch Biopsy [96].

During the shave biopsy the portion of the lesion that is above the level of the surrounding skin is shaved off using a blade. Superficial lesions such as seborrheic keratosis can be biopsied in this manner. However, it is better avoided as it does not include deeper tissues.

Saucerization biopsy is ideal for vesiculobullous disorders and also for epidermal neoplasms like seborrheic keratosis. In this technique, the plane of cleavage passes through the reticular dermis and occasionally through the subcutaneous tissue [94]. This is performed using a shaving blade. After anaesthetizing the area, the shaving blade is held between the thumb and the index finger such that the blade is bent to form an arc. The blade is introduced through the skin such that it goes through the dermis. The lesion is shaved completely in this manner and the wound is dressed to heal by secondary intention.

Wedge biopsy is usually performed for large lesions where not only length and breadth but also depth of specimen is important [94]. Hence, this technique is commonly chosen for subcutaneous mycosis, margin of squamous cell carcinoma, tuberculosis verrucosa cutis, etc. After anaesthetizing the skin, a stab incision in a "*V*" or a triangular shape is performed with scalpel blade so as to extract a cone of tissue. Most often the lesion is left to heal by secondary intention.

The results of the histopathological analysis heavily depend on the quality of the biopsy submitted to the specialist. Frequently, a poorly performed skin biopsy may cause the specialist to make an incorrect diagnosis. The clinical information provided to the histopathologist also influences the final diagnosis. To increase the accuracy of the diagnosis, samples are often sent for analysis to two or more specialists. However, even if the diagnosis does not need to be confirmed by other specialists, a skin biopsy (including histopathological analysis) is a rather time-consuming procedure that is also at risk of human error. All this makes the task of developing new non-invasive methods for diagnosing different types of skin cancer suitable for mass screening quite urgent.

1.2.3.3. Histological analysis

The final stage of the standard clinical procedure for diagnosing a skin lesion is the histological analysis of its tissue. For this purpose, tissue samples obtained after biopsy are placed in formalin to preserve them during transport.

After obtaining tissue samples, the histopathologists proceed to the formation of slices, for which the tissue sample is first placed in liquid paraffin, which hardens at room temperature. Fixation of samples in paraffin allows to obtain the tissue slices with a few micrometres thick. Then follows the staining of slices with various dyes, which allows to highlight the features of the structure of tissues and cells. This stage provides the possibility of more detailed analysis of morphological characteristics of the studied samples under the microscope. Based on the results of microscopic analysis of the slices, specialists can make an accurate diagnosis and the choice of treatment methods.

1.2.4. Present limitations and needs for clinicians

Despite the important advantages of the standard diagnostic procedure, such as a high Recall value (as almost all suspicious lesions are removed) and a relatively high correctness of diagnosis (after histological analysis), it also has a number of limitations and disadvantages. For example, it is virtually impossible to determine the exact type of tumour and the extent of its spread by visual inspection, requiring biopsy and histological analysis that significantly increase the time before the final diagnosis.

The procedure of taking a biopsy is quite painful and invasive. In addition, in the case of excisional biopsy, an accurate assessment of the excision margins of the tumour is required, as an erroneous assessment of safe margins may result in the need to repeat the biopsy. A possible solution to this problem is to perform histological examination at the time of surgery

(Mohs micrographic surgery, extemporaneous histological examination), however, even then, the entire procedure can take several hours and be extremely uncomfortable for the patient.

It is also important to notice that biopsies are often used for any skin neoplasms that are suspected to be skin cancer. Thus, the ratio of biopsies to skin lesions has increased from 1.134 in 1993 to 2.058 in 2016 [49].

Histological analysis, in its turn, requires careful handling of specimens and certain equipment, which makes the diagnostic procedure even more expensive and time-consuming.

Thus, the development of new non-invasive methods that could be used both to determine the type of skin lesion and to assess its boundaries is of great interest. A promising solution to this problem is the use of optical methods.

1.3. *In vivo* skin Optical biopsy

. Optical methods of skin tissue *in vivo* characterization (also referred to as "Optical Biopsy") are the subject of numerous developments and investigations for more than two decades. Indeed, they can be a good solution that fulfils all of the criteria aforementioned (*in vivo*, non invasive, real time, Se, Sp), as they are able to probe the modifications of the optical properties of the skin related due to the physiological and metabolic changes along with the development of cancer and precancerous pathologies.

1.3.1. Changes in skin optical properties related to the development of carcinomas

As described in the previous section, cancerous and precancerous skin conditions have a significant effect on the structure and the metabolism of l skin components at multiple scales (sub-cellular to tissue). As a consequence, the optical properties of the skin, such as absorption, scattering or intrinsic fluorescence, are significantly modified during carcinogenesis. Changes in the scattering properties of the skin are usually associated with structural changes in different layers of the skin, such as changes in cell size and shape, cell arrangement pattern and nucleus/cell ratio. The absorption properties (and consequently the fluorescence) of the skin depend mainly on the composition and concentrations of light absorbers in the measured sample, such as collagen, elastin or haemoglobin, the concentration of which also depends on changes in the structure of blood vessels in the lesional area.

Changes in skin properties and structure due to carcinogenesis occur in all affected skin layers. For example, for the epidermis, a focal, atypical honeycomb pattern related to chronic

sun exposure is observed in actinic keratosis, while a disarranged epidermal pattern and ulceration are commonly observed in SCC. At the dermal level, dilated vessels may be observed in NMSC, where they are typically horizontal and branching, surrounding the tumor island in BCC and more irregular in SCC. Some other morphological variations at the dermal level include bundles of lace-like collagen in actinic keratosis and SCC [97] and collagen bundles surrounding dark silhouette in hypopigmented BCC [98].

It is important to note that optical properties also depend directly on the photon energy and, therefore, on the wavelength. Thus, the choice of wavelength range and experimental research method should be based on the changes in optical properties of biological tissue components in this range and the possibility of detecting these changes using an appropriate method. For example, the biophotonic features of collagen, elastin, keratin, melanin, NADH and some other skin components are well studied in the visible and infrared (IR) ranges, which, considering changes in the concentrations of these components during carcinogenesis, allows using such methods as Raman spectroscopy or fluorescence spectroscopy for skin cancer diagnostics. At the same time, the multilayer multicomponent structure of the skin makes the interpretation of the obtained spectra rather complicated, as they contain a mixed signal obtained from all skin layers and different structures in this layers. An example of absorption and extinction (sum of absorption and scattering) coefficient spectra of some skin components in the visible and near-infrared range is shown in **Figure 1.5**. It can be seen that even for the presented components the absorption spectra may overlap, which makes the interpretation of the obtained results a rather complicated task.

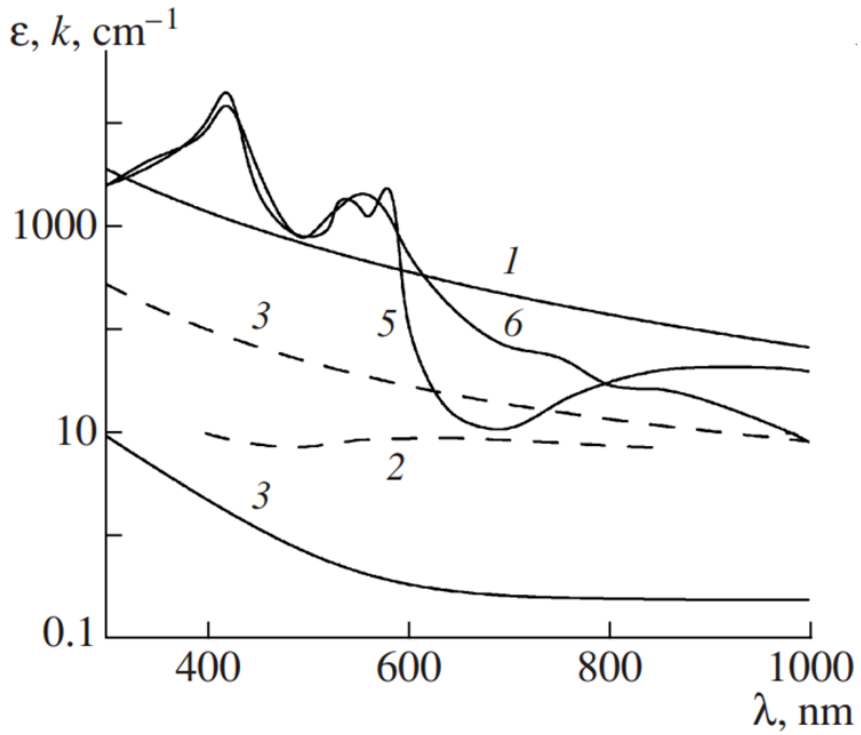


Figure 1.5. Absorption index (solid curves) and transport extinction index (dashed curves) of (1) melanin, (2) blood, (3) basic tissue, (5) oxyhemoglobin, and (6) deoxyhemoglobin [99].

There are a large number of optical parameters that can be determined for biological tissues [100], however, almost all of them cannot be measured *in vivo*, which makes them difficult to use for non-invasive diagnostic tasks.

In most studies dealing with the determination of the optical properties of skin, the main focus is placed on the determination of absorption and scattering coefficients, characterizing the attenuation of light intensity in the tissue [101–103]. These parameters are sensitive to changes in the structural and functional properties of the skin, and therefore can be informative for the diagnosis of skin diseases, including skin cancer. However, according to [104] and [105], the use of such parameters for skin cancer diagnosis in the visible and near infrared range is not possible because the differences in the properties of healthy and cancerous tissues are not statistically significant, which may be a consequence of the large variability of skin properties depending on the anatomical site, phototype, as well as the age and sex of the patient.

Thus, the use of any one optical method or optical parameter alone may not be sufficient for accurate diagnosis of NMSC due to complex skin structure and composition. At the same time, it is of great interest to combine several complementary methods in one study, which

allows to use a wider list of parameters for diagnosis that carry information about changes in both skin structure and composition. Such studies are called multimodal studies and will be described in more detail in **Section 1.3.3**.

1.3.2. Single modality studies

1.3.2.1. Spectroscopic methods

Optical biopsy methods can be roughly divided into spectroscopic and imaging methods. The basis of spectroscopic methods is the measurement of intensity spectra which is the dependence of one or more optical parameters (reflection, scattering, radiative and/or non-radiative absorption) of a biological tissue on frequency or wavelength under specific conditions. As it was mentioned earlier, the spectral shape of these spectra is quite sensitive to changes in skin properties, therefore, spectroscopic methods can be used to diagnose various pathologies, including skin cancer. However, for each method it is necessary to choose the parameters of the source of radiation (wavelength, power, spatial resolution, etc…) as well as the wavelength range of the measured spectra. In the case of skin cancers, numerous studies reporting clinical tests of spectroscopic techniques such as fluorescence spectroscopy, diffuse reflectance spectroscopy and Raman spectroscopy are considered in the next state-of-the art paragraphs.

Fluorescence spectroscopy

Light-induced Fluorescence Spectroscopy or Intrinsic/Auto-Fluorescence spectroscopy (AFS) is a sensitive tool whose accuracy in diagnosing skin cancer or other skin pathologies depends on many factors, such as pigmentation, morphology, localisation and growth stage of the lesions under investigation. Skin intrinsic fluorescence spectra contain information about all these properties, which makes AFS a method of primary interest for diagnosing various skin pathologies, including different types of skin cancers [10, 106].

The principle of AFS is based on the excitation of fluorescent molecules naturally present in the biological tissues(skin fluorophores) capable of re-emitting light after the absorption of photons of a certain energy. The excitation energy of fluorescence differs between fluorophores, thus requiring the use of lasers with different wavelengths.

According to [107], the most contributing cutaneous fluorophores, such as collagen, elastin, keratin and their cross-links, NADH and porphyrins, have fluorescence excitation peaks in the 320-420 nm wavelength range, with the consequence that most studies on the use of AFS

for skin cancer diagnosis use narrow band or monochromatic light sources in this wavelength range.

When fluorescence from multiple skin fluorophores is simultaneously excited, the resulting spectrum will contain contributions from each fluorophore. This approach can help to increase the amount of information in the spectrum useful for classification, and thus increase the resulting diagnostic accuracy. At the same time, the fluorescence signal from different fluorophores may overlap, which can negatively affect classification accuracy. In addition, the interpretation of the results is considerably complicated in any case. An example of AF spectra measured on normal skin, benign lesions, BCC and SCC is presented in **Figure 1.6** [16]. It can be seen that the presented spectra of cancerous lesions are visually different from those of normal skin and benign lesions. However, a deeper interpretation of the results is almost impossible. Thus, the use of machine learning methods is usually necessary to separate diagnostic classes based on AF spectra [18].

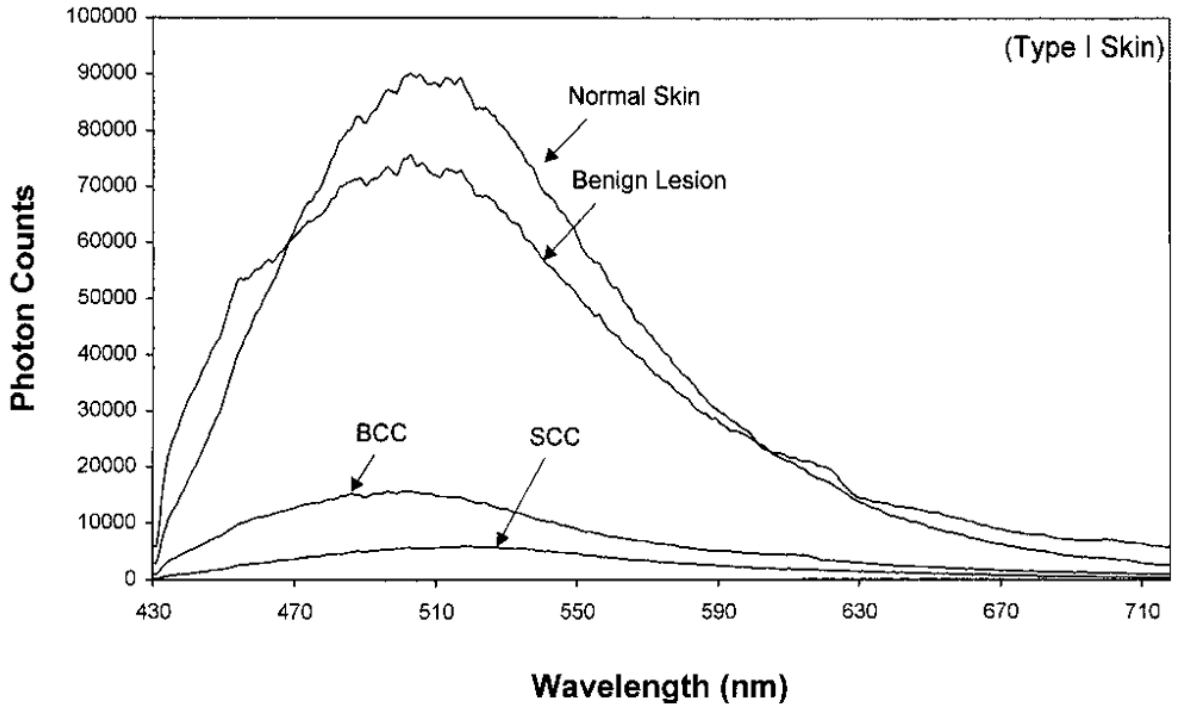


Figure 1.6. Example of AF spectra of BCC, SCC, heathy skin and benign lesion [16].

The interpretation problem can be partially solved by using multiple sources of fluorescence excitation with different wavelengths, which makes it possible to vary the contribution of different fluorophores to the resulting spectra. Such an approach is used in [17], where 4 different lasers in the spectral range of 340-850 nm were used to induce autofluorescence for the diagnosis of BCC. The dataset consisted of 286 skin samples, 137 of

which were identified as cancerous. The sensitivity and specificity obtained using a deep learning feed-forward neural network were 62% and 83% respectively.

At the same time, melanin absorption also has a significant effect on the shape of the autofluorescence spectra [108], as a consequence of which the accuracy of NMSC diagnosis using AFS may vary for different skin phototypes. In [16], AFS with an excitation wavelength peak at 410 nm and a spectral range of fluorescence acquisition from 430 to 716 nm was used for the diagnosis of BCC, SCC, benign and precancerous lesions for different types of skin (I, II and III). The maximum accuracy of the classification of cancers obtained was 93% for skin type I, whereas for skin type II and III it was 89% and 78%, respectively. Benign and precancerous lesions were classified with an accuracy of 100%, 46% and 27% in patients with skin types I, II and III, respectively.

Besides AFS, it's worth mentioning that, in some specific cases of skin cancer diagnosis and treatment, exogenous fluorophores or exogenously-induced fluorophores can be used for photodetection and photodynamic therapy. Exogenous fluorophores are used when the skin lesion is heavily pigmented. In such cases, the fluorescence signal is often too weak to be used for diagnostic purposes. Hence, exogenous fluorescent agents used to enhance the fluorescence signal from skin lesions are various forms of photosensitizers - drugs that selectively accumulate in cancer cells and fluoresce under the influence of photons of a particular wavelength. In addition, they produce a photodynamic effect that can be used for therapeutic treatment of the examined tumours. Thus, diagnosis and therapy can be performed in a single procedure, which could be very useful for clinical practice [10, 109].

However, the combined use of exogenous fluorophores for diagnosis and therapy is only advisable if the specialist has reliable information that the skin lesion is malignant. The most reliable method of diagnosis in this case is still the standard dermatological examination and surgical biopsy. Therefore, the use of exogenous fluorophores during the initial diagnosis of a suspicious lesion, which may be malignant, dysplastic or benign, is clinically non-optimal [107] due to: (i) a significant complication of the diagnostic procedure, (ii) the increased time required for the procedure, as accumulation of the drug in the skin lesion under examination can take up to several hours, (iii) the necessity of an invasive application of the drug and (iv) low diagnostic specificity.

Diffuse reflectance spectroscopy

Diffuse reflectance spectroscopy (DRS) is another spectrally-resolved method that can be used for rapid and non-invasive diagnosis of skin diseases. The DRS spectrum of the skin mainly depends on scattering and absorption properties over a wide wavelength range. Therefore, diffuse reflectance spectra obtained on skin contained information on the morphology of normal or diseased cells and tissues as well as on differences in concentrations of different endogenous skin components (e.g. haemoglobin, melanin, bilirubin and water) which have absorption bands in the corresponding range. The use of DRS as a diagnostic tool is based on the fact that many pathologies cause significant structural changes at the cellular and subcellular levels that have a significant effect on the spectrum of diffuse reflected light [6, 10].

The principle of DRS is the illumination of a sample using a beam of light in a wide wavelength range and the detection of the reflected light. The scattering centres beneath the skin surface also have a significant influence on the spectrum. Analysis of differences in diffuse reflectance spectra of normal and abnormal tissue allows researchers to obtain information about changes in the geometry of internal scattering centres and, therefore, the corresponding skin components [110]. In addition to scattering properties, the shape of DR spectra is also significantly influenced by the absorption properties of haemoglobin and melanin [111, 112].

Therefore, the DRS technique applied to the diagnosis of skin pre-malignant and malignant pathologies has been studied for more than two decades because it provides DR spectra containing information both about the structure and the chemical composition of the tissue under study. In addition, the instrumental implementation of DRS is relatively simple, and as a result, this method is often used in research for the diagnosis of various diseases. The main drawback of this approach is the difficulty of interpreting the results, which requires the use of modern numerical methods. Examples of DR spectra of cancerous and benign skin tissues are shown in **Figure 1.7** [113].

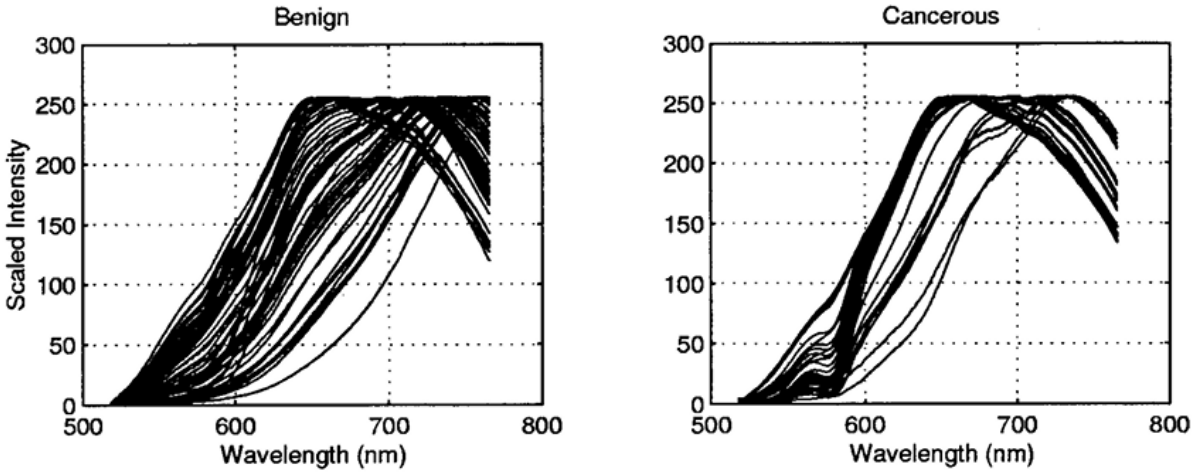


Figure 1.7. Example of DR spectra of cancerous and benign tissues [113].

In most studies focussing on the use of DRS for the diagnosis of NMSC, the visible range was used, and the resulting sensitivity and specificity values exceeded 80% . In addition, machine learning was used to classify lesions in almost all the studies found, which also indirectly confirms the complexity of analysing DR spectra.

Raman spectroscopy

Raman spectroscopy (RS) is a method based on the inelastic scattering of monochromatic light (usually in the visible or near-infrared range) which is used to analyse the vibrational modes of molecules. The change in energy of photons due to their interaction with the substance correlates with the difference between the initial and final vibrational energy levels of the scattering molecule, providing information about the type of molecule and its photon mode in the sample. In a Raman spectrum these variations appear as a series of peaks. The positions and relative magnitudes of these peaks correspond to vibrational energies associated with particular chemical bonds in particular molecules. Many molecules have distinguishable spectra, so the Raman spectrum can be used to determine the molecular composition of a sample with high accuracy [114].

The main advantage of Raman spectroscopy is the high interpretability of the results and its main drawback is the low values of inelastic light scattering cross sections, which leads to longer integration time, that makes *in vivo* studies more difficult for patients and medical doctors. Moreover, the more complex design of the device to obtain acceptable values of signal-to-noise ratio is another drawback [115]. In addition, the depth of photon detection for RS (from tens to hundreds of micrometres) is quite small compared to AFS and DRS (up to 1-2 cm) [116].

Despite these drawbacks, RS remains a powerful diagnostic tool for a wide variety of pathologies, including NMSC.

At the same time, despite the relatively high interpretability of the results compared to AFS and DRS, the separation of skin spectra obtained using RS can be a complex task requiring the use of statistical or ML techniques. The examples of Raman spectra measured in different types of cutaneous lesions are presented in **Figure 1.8** [117].

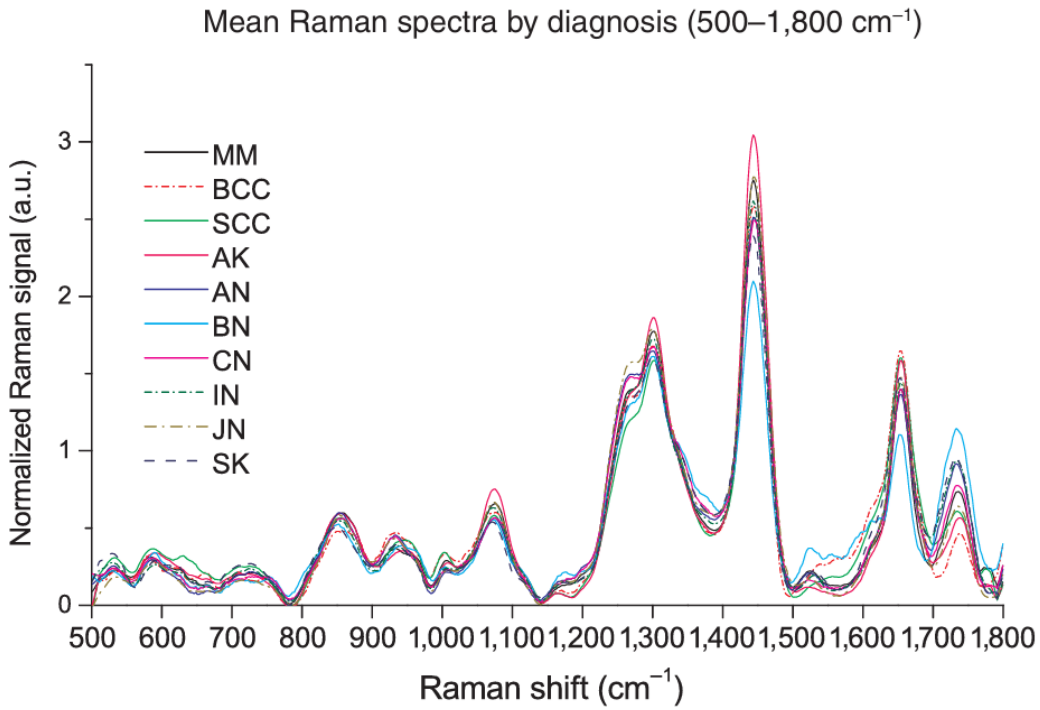


Figure 1.8. Example of Raman spectra of cutaneous lesions of different types [117].

There are various studies on the use of RS for NMSC diagnostics both in *ex vivo* [20, 38, 118] and *in vivo* [20, 22, 23, 119, 120] format. Sensitivity and specificity in most cases also exceed 80%.

A summary of the results of the search for articles on NMSC diagnostics using Autofluorescence Spectroscopy, Diffuse Reflectance Spectroscopy and Raman Spectroscopy is presented in **Table 1.2**.

Table 1.2. Summary of the results of the search for articles on NMSC diagnostics using spectroscopic methods.

Reference	Sample types	Number of samples	Acc/Se/Sp (%)
Autofluorescence Spectroscopy			
[16]	BCC	88	92/-/- (skin phototype I) 79/-/- (skin phototype II) 53/-/- (skin phototype III)
	SCC	13	
	Pre-cancerous lesion	12	
	Benign lesion	47	
	Healthy	119	
[17]	BCC	137	73/62/83
	H	149	
[18]	BCC (All types)	83	-/94/94 (multiclass mode) -/93/96 (BCC vs Psoriasis)
	Bowen's Disease	4	
	Psoriasis	37	
Diffuse Reflectance Spectroscopy			
[13]	BCC/SCC	127	92/92/92
	AK/SK	50	
[14]	BCC	19	-/-/-
	SCC	38	
	AK	14	
[15]	BCC	59	78/78/80
	Normal	125	
[121]	BCC	9	86/87/85
	SCC	4	
	MSC	1	
	Benign	13	
[122]	BCC	10	93/90/94 (BCC vs MSC+SCC) 82/83/80 (MSC vs SCC)
	SCC	5	
	MSC	12	
Raman Spectroscopy			
[20]	BCC	28	94/94/94 (for both Normal vs BCC vs SCC+AK and Normal vs BCC+SCC vs AK)
	SCC	7	
	AK	11	
	Normal	10	
[22]	Cancer/precancer	340	-/90/74 using spectra only -/90/80 with demographics
	Benign	391	
[23]	BCC	50	-/90/92
	Normal	173	
[119]	BCC	9	95/100/91
	SCC	4	
	Scar	8	
	Normal	21	
[120]	BCC	28	78/-/- (multiclass mode) 92/94/89 (Normal vs Rest)
	SCC	7	
	AK	11	
	Normal	53	

1.3.2.2. Imaging methods

Unlike spectroscopic methods, imaging methods make it possible to obtain images of the surface of the skin or the structure of its layers in depth. As a result, these methods are also well suited for the diagnosis of skin cancer. The most widespread research is on the use of reflectance confocal microscopy, fluorescence lifetime imaging microscopy, multiphoton microscopy, optical coherence tomography and multi-hyper-spectral imaging. Examples of skin images obtained using the methods listed above are shown in **Figures 1.9.** and **1.10**.

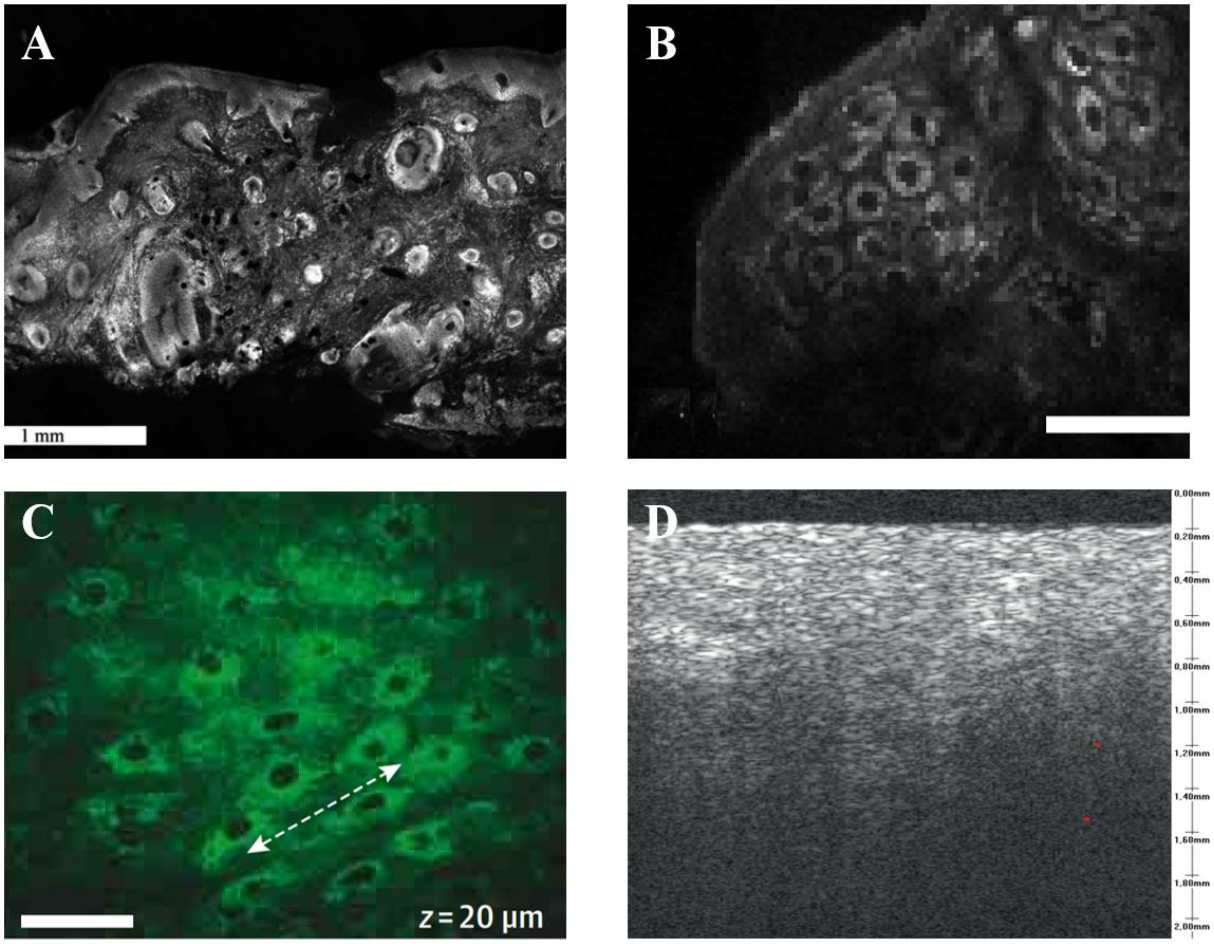


Figure 1.9. Images of human skin obtained from BCC lesion and healthy skin using Reflectance Confocal Microscopy (A) [123], Fluoresce Lifetime Imaging Microscopy (B) [124], Multiphoton Microscopy (C) [125] and Optical Coherence Tomography (D) [126].

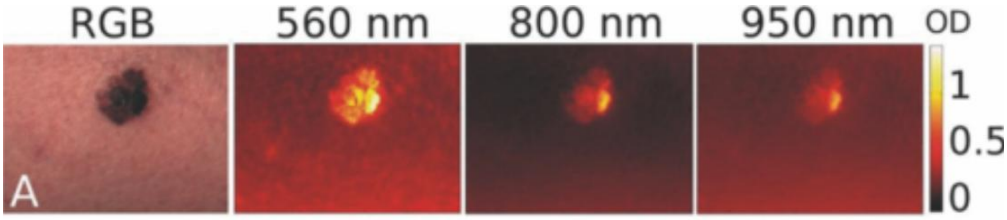


Figure 1.10. Images of human skin obtained from MM and using Multispectral Imaging [127].

Reflectance Confocal Microscopy (RCM)

Reflectance Confocal microscopy (RCM) is an optical imaging technique for non-invasive tissue imaging with higher resolution and contrast than conventional microscopy. These advantages are obtained with the use of a point light source to illuminate the specimen and by placing a hole between the tissue specimen and the detector, which rejects the multiple scattered out-of-focus light, allowing only the out-of-focus light from the specimen to reach the detector. This image is an optical slice reflecting a single focal plane in the sample under study. By changing the focal plane or moving the sample, a series of images can be obtained at different positions throughout the thickness of the sample, and by using optical sectioning, a three-dimensional image of the sample can be obtained [128, 129].

RCM has been used for the diagnosis of NMSC [123] and melanoma [130–133] as well as for the diagnosis of precancerous conditions [134]. The sensitivity and specificity values were also up to 97.5% and 99%, respectively. In all cases, the data were manually classified based on typical criteria for standard biopsies. RCM has a shallower depth of penetration than OCT, but slightly better spatial resolution, as well as high contrast.

Fluorescence Lifetime Imaging Microscopy (FLIM)

The basic idea of Fluorescence Lifetime Imaging Microscopy is to excite the fluorescence of a sample and measure the time interval during which the fluorescence intensity decreases by a factor of *e* relative to its peak value. This time interval is called lifetime and has different values for different agents, which allows to obtain information about changes in the concentration of many different components of the sample under study. Unlike Fluorescence spectroscopy, FLIM does not deal with fluorescence intensity, which makes it more suitable for analysing samples containing several fluorophores with close emission spectra [135]. FLIM has been successfully applied for studies on BCC [124, 136, 137], SCC [138] and MSC [135, 139] diagnostics.

Multiphoton Microscopy (MPM)

In fluorescence spectroscopy and microscopy, excitation of fluorophores occurs by absorbing a single and relatively high energy photon in the visible or ultraviolet range, as contrast to MPM where excitation of fluorophores is based on the simultaneous absorption of two or more photons of lower energy, typically in the near infrared region of the spectrum. For efficient excitation of fluorophores by this method, extremely short laser sources with high peak power and pulse duration in the range of tens to hundreds of femtoseconds must be used [140].

With MPM, only a very small volume of fluorophore is excited at the focal point, enabling optical sectioning with subcellular resolution without the need for a confocal pinhole, so that MPM can be used for imaging highly scattering tissues [140].

In addition to multiphoton excitation of fluorescence, the ultrashort femtosecond laser is also effective for obtaining the nonlinear polarisation effect of second harmonic generation (SHG) as a result of interaction with non-centrosymmetric biological structures such as collagen and myosin [140–142]. Photons generated as a result of SHG have a wavelength being exactly the half of the wavelength of the incident photon, as a consequence of which SHG signals can be separated from fluorescence quite easily and provide contrast in the tissue imaging. The structural sensitivity of SHG can also be exploited to detect collagen and myosin structural transitions in normal and pathological conditions with higher sensitivity than that achievable for conventional histology [140].

There are a few studies on the use of MPM for imaging different types of skin cancers [125, 140, 143–145], but the number is smaller than for RCM or OCT, which may be related to the fact that the devices needed for MPM are not widely available in medical and research institutions compared to RCM and OCT.

Optical Coherence Tomography (OCT)

Optical coherence tomography is an interferometric method based on the registration of scattering of probing radiation on optical inhomogeneities of objects under study with measurement of the corresponding time delay and construction of a 2D or 3D images based on the obtained data. OCT methods are commonly used to study the internal structure of biological tissues in the near-IR wavelength range (800 - 1300 nm). This wavelength interval is also called the "therapeutic transparency window" because light scattering in this range greatly exceeds absorption. When a bio-object is illuminated by near-IR radiation, the light penetrates into the

medium to a depth of several millimetres, but most often data analysis for lesion classification is performed for depth values not exceeding 1 mm.

There are different types of OCT systems. The first version of OCT was Time-domain OCT (TD-OCT), the main technical feature of which was the fast translation of the reference mirror of the internal interferometer, allowing for the axial time-dependent scanning of skin [26]. Full-field OCT (FF-OCT) was developed based on TD-OCT. FF-OCT is characterized by the ability to acquire "in face" skin images without the need to scan along the specimen [174]. Another method is Frequency- (or Fourier) domain OCT (FD-OCT). Its feature is the absence of scanning of the reference mirror; instead, scanning is carried out by the wavelengths of the source [175]. Finally, a recently invented technique is Line-field Confocal OCT (LC-OCT). It combines the advantages of TD-OCT and confocal microscopy and provides 3D images of tissue at high-resolution allowing to distinguish the cellular structure of the skin [176].

OCT has been used for diagnosis of NMSC [126, 146–148], melanoma [149, 150], precancerous lesions [151], and for evaluation of tumour borders [152, 153] with sensitivity and specificity up to 95%. The advantages of this technology include high spatial resolution, as well as greater (compared to RCM) depth of sample sensing. The main drawback of OCT in the context of its use for skin cancer diagnosis include high cost, and the lack of biochemical information about the imaged skin layers.

Multi-spectral and Hyper-spectral Imaging

Multi-spectral imaging (MSI) is another imaging modality increasingly investigated over the last decade by researchers in dermatology. MSI systems allow to image reflected or transmitted light from the skin with high spatial resolution and in several spectral bandwidths. Thus, this modality provides appropriate spectral resolution of interest to identify the contribution of main chromophores such as melanin, hemoglobin, water, beta-carotene, collagen and bilirubin. [154]. The main difference between Hyper-spectral imaging (HSI) and MSI lies in the number of spectral channels (**Figure 1.11**). HSI uses almost continuous and contiguous ranges of wavelengths (e.g. wavelength range 400 - 1100 nm in steps of 1 nm) whilst MSI uses a subset of targeted wavelengths at chosen locations (e.g. in 400 - 1100 nm in steps of 20 nm).

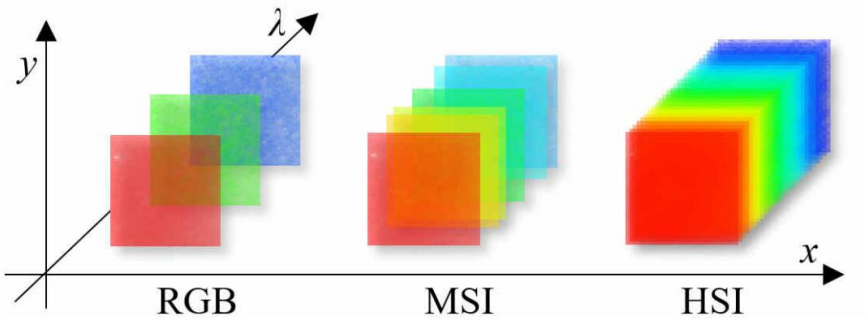


Figure 1.11. Difference between RGB imaging, multispectral imaging and hyperspectral imaging [155].

MSI systems can sequentially illuminate the skin in different spectral ranges and collect the reflected light on the sensor. They can also sequentially filter the reflected or transmitted light coming from the skin before it reaches the camera sensor. These two technological approaches (active and passive) for light collection give systems pros and cons for their applications in skin analysis: active illumination systems are typically based on light emitting diodes (LEDs), which are low cost, small in size, long-lasting and low power consumption. This approach makes optical designs of active systems more compact than that passive systems. However, passive systems provide continuous spectral sampling because they are based on electro-optical devices such as liquid crystal or acousto-optic tunable filters [154, 156]. The spectral range can be either at the visible wavelengths (VIS, from 380 to 780 nm) or the near-infrared range (NIR, from 780 to 2500 nm).

The data provided by MSI systems are referred to as spectral cube, which contains all spatial and spectral data of the analysed sample. Depending on how the spectral cube is constructed, MSI devices are divided into different types [12]. All MSI systems designed for skin analysis belong to the type of area scanning systems or snapshot imaging systems. The use of MSI for melanoma diagnosis has been described in [157–159] and melanoma and AK in [160]. The main advantages of using MSI as a skin diagnostic tool include the possibility of imaging in a wide spectral range (although smaller than that of HSI). The main limitation is the inability to take into account the contribution of internal skin layers in the resulting spectra, as well as the loss of some information (in comparison to HSI) which makes analysing and interpreting the results quite difficult. HSI as a diagnostic tool for skin cancer is described in [161, 162] for melanoma, in [163] for NMSC and in [164] for NMSC and melanoma together. A summary of the results of the search for articles on NMSC diagnostics using Imaging methods is presented in **Table 1.3**.

Table 1.3. Summary of the results of the search for articles on NMSC diagnostics using imaging methods.

Reference	Sample types	Number of samples	Acc/Se/Sp (%)
Reflectance Confocal Microscopy			
[123]	BCC	48	-/97/89
Fluorescence Lifetime Imaging Microscopy			
[124, 136]	BCC	3	-/-/-
	Naevus	4	
[137]	BCC	3	-/-/-
	AK	3	
	Bowen disease	3	
Multiphoton Microscopy			
[125]	BCC	10	-/-/-
[143]	BCC	9	-/-/-
	SCC	6	
	Normal	14	
Optical Coherence Tomography			
[126]	BCC	34	-/-/-
[146]	BCC	37	77/-/-
	AK	41	
[147]	BCC	3	-/79-94/85-96 (Normal vs Lesions) -/60-100/60-91 (BCC/SCC vs Normal)
	AK	4	
	Benign lesions (BL)	4	
	Normal skin	20	
Multi-spectral and Hyper-spectral Imaging			
[163]	BCC	67	-/-/-
	SCC	22	
	AK	5	
[164]	BCC	19	-/63/72 (MSC vs BCC + BL) -/91/91 (All Cancers vs Normal)
	Benign lesions (BL)	10	
	MSC	16	
	Normal	45	

1.3.3. Multimodal approaches

Most of the methods described in the two previous sections focus on a particular skin property, or group of properties, while information about other parameters is either ignored or used incompletely. Thus, only part of the full diagnostic potential of each method is inevitably exploited by each method. There are two main ways to solve this problem. The first way is to use several variations of a single approach or to combine several different approaches in a single method. Such methods are referred to as multimodal. The other way is a deeper and more comprehensive analysis of the obtained information, which will be described in detail in the second chapter. Another important aspect discussed in the second chapter will be the correct combination of information obtained from multiple modalities.

Multimodal devices should provide the opportunity to measure samples at the same points based on several different experimental methods. As a result, the creation of such devices is associated with many difficulties, since each optical method has a certain set of components that are not always possible to combine within the framework of one device and one optical probe. One of the solutions to this problem is the use of fiber-optic probes, which, moreover, can be used to measure different parts of the body due to their small size. However, the use of such a configuration also significantly limits the list of methods that can be successfully combined in one device. As a result, almost all of the studies found on the use of multimodal spectroscopic systems for the diagnosis of skin cancer are based on various combinations of AFS, DRS and RS.

AFS and DRS provide complementary information about the absorbing and dispersing properties of the skin, which makes their use in multimodal systems extremely relevant. RS, in turn, has a high interpretability, which is extremely important in a more detailed analysis of the results obtained. At the same time, each multimodal system is unique, as a result of which it will be analysed separately.

There are also several studies related to the use of multimodal imaging techniques for the diagnosis of NMSC [165–169], however, the modalities used in some of these studies are not complimentary to the changes in the optical properties of the skin that they can register. Furthermore, due to the fact that the topic of this thesis is related to spectroscopic multimodal methods, studies on the use of imaging modalities will not be discussed in detail.

In [170], a combination of AFS with an excitation wavelength of 340 nm and DRS, with a halogen lamp light source with an emission range of 400-900 nm, was used. The data set of 24 patients contained benign (dermal nevi, compound nevi, dysplastic nevi) and malignant (melanoma) lesions. When the methods were used separately in the diagnosis of melanoma, sensitivity (77.8%) and specificity (93.3%) were similar for all methods. The best results (sensitivity = 100%, specificity = 93.3%) in this study were achieved for the combination of AFS and DRS, which shows the greater efficiency of multimodal systems compared to unimodal systems.

In [35, 171] a similar method of skin lesions analysis was also applied; however, AF spectra were acquired for excitation wavelength peaks of 365 nm, 385 nm and 405 nm. The parameters of the source for obtaining DR spectra were unchanged. Another important change is that the data set under study included more than 400 patients whose pathologies were

classified as BCC, SCC, melanoma, and cutaneous dysplastic and benign cutaneous lesions. Samples were compared in three bi-class classification ways: NMSC vs. benign, MM vs. benign, and MM vs. NMSC. Sensitivity and specificity were 100% and 99.2% in the first two cases, and 99.1% and 90.5%, respectively.

In [172], a combination of AFS with an excitation wavelength of 445 nm and DRS with a wavelength range of 350-700 nm was used for the diagnosis of NMSC (BCC and SCC) and AK. Classification was performed in three ways: normal vs BCC, normal vs AK and SCC, and AK vs SCC. The highest values of sensitivity and specificity in all cases were 94% and 89%, 97% and 62%, and 100% and 50%, respectively, but only in the latter case did the combined approach (DRS + AFS) show the best results. In other cases, the use of DRS spectra provided the highest classification accuracy.

In [34], a combination of DRS (four sources: LED with emission wavelength 430 nm, white light LED with 450-700 nm range, 750 nm and 850 nm near-infrared LEDs) and time-resolved AFS with three excitation wavelengths (355, 375, and 395 nm) was used to diagnose BCC. The resulting sensitivity and specificity values were 100% and 71%, respectively.

Another quite common combination of methods is the combination of RS and AFS. For example, RS was used in the study of amelanotic melanoma, pigmented melanoma, and BCC [37, 173, 174] with an excitation wavelength of 785 nm and an imaging range of 1800-1200 cm^{-1}. Therefore, analysis of the spectrum in the entire spectral range allowed considering both AF and Raman spectral characteristics of the tissue under study. The best accuracy obtained for discriminating all melanomas and BCC was 90%, and separation of two types of melanomas was 82 %.

In [33], RS with an excitation wavelength of 830 nm and a spectral range of 1796 to 840 cm^{-1}, AFS with an excitation wavelength of 337 nm and DRS with a spectral range from 330 to 690 nm were combined in one device. The dataset contained both NMSC and melanoma samples, AK and pigmented lesions. The resulting sensitivity/specificity values were 100%/100 % for melanoma vs pigmented lesions and melanoma and pigmented lesions vs normal skin, 95%/71% for SCC and BCC vs AK and 90%/85% for AK and NMSC vs normal skin. All analyses were also performed separately for each modality, but the combined approach showed similar or better results in all cases. A similar approach is presented in [175]. A summary of the results of the search for articles on NMSC diagnostics using Autofluorescence Spectroscopy, Diffuse Reflectance Spectroscopy and Raman Spectroscopy is presented in **Table 1.4**.

Table 1.4. Summary of the results of the search for articles on NMSC diagnostics using multimodal methods.

Reference	Sample types	Number of samples	Acc/Se/Sp (%)
Imaging methods			
[165]	Nodular BCC	17	-/-/-
	Infiltrative BCC	10	
	Invasive SCC	10	
[167]	BCC	5	-/-/-
	SCC	5	
	Normal	14	
[168]	Benign	63	97/98/95
	BCC	53	
	SCC		
	MSC		
[169]	BCC	8	-/-/-
	Normal		
AFS and DRS			
[34]	BCC	10	-/100/71 (BCC vs Normal)
	SCC	3	
	AK	3	
	Other lesions	11	
	Normal	27	
[35, 171]	BCC	143	-/100/99 (BCC + SCC vs Benign) -/99/90 (MSC vs Benign)
	SCC	48	
	MSC	62	
	Benign lesions		
	Dysplastic lesions		
[172]	BCC	18	-/94/89 (BCC vs Normal) -/97/62 (AK + SCC vs Normal)
	SCC	34	
	AK	39	
	Normal	56	
AFS and RS			
[37]	BCC	40	-/100/96 (BCC vs MSC)
	Melanoma	39	
	Normal	79	
[174]	BCC	120	90/-/- (BCC vs MSC)
	Melanoma	69	
AFS, DRS and RS			
[33]	BCC and SCC	57	90/-/- (SCC + BCC vs AK) 87/-/- (AK + SCC + BCC vs Normal)
	MSC	12	
	AK	14	
	Pigmented lesions	17	
	Normal	71	
[175]	BCC	5	100/100/100 for both BCC vs Normal and SCC vs Normal
	SCC	5	
	Normal	5	

Multimodal devices developed in CRAN

It can be understood from previous paragraph that most of the listed multimodal methods were based on combining AFS with several excitation wavelengths and DRS using optical fibre probes. As mentioned above, such multimodal methods provide complementary information about changes in optical properties of the skin, which makes the development and implementation of appropriate devices an extremely promising task.

Two devices combining AFS with different excitation wavelengths and DRS were developed in our team. The first device did not have a commercial name and will be referred to in this thesis as the Pilot device. The pilot device combined the ability to measure AF spectra for seven different excitation wavelengths in the 360-430 nm range and DR spectra in the visible range. A xenon lamp with an emission range of 300-800 nm was used as a light source. A system of linearly variable band-pass filters was used to obtain narrow bands of autofluorescence excitation. The configuration of the multiple fibre optic probe allowed all spectra to be acquired for several different values of source-detector separation (SDS), which is a very significant aspect, as increasing the distance between the source and the detector, the proportion of photons with a greater penetration depth into the skin is also increased, which makes it possible to obtain information about the condition of deeper layers of the skin. The optical probe of the device allowed spectra to be measured for three SDS values: SDS_1 = 271μm, SDS_2 = 536μm and SDS_3 = 834μm. The pilot device was used in the series of preclinical studies to collect multiply excited AF and DR spectra from the skin of mice with light-induced precancerous and benign skin lesions [42, 43, 46, 47]. The results of the latest study [47] will be described in the third chapter of this thesis (first main contribution).

The next multimodal device developed by our team was the SpectroLive Medical Device, which also provided the ability to capture AF and DR spectra in the visible range [40]. Five LEDs with excitation wavelengths in the range of 365-415 nm were used to obtain skin AF spectra. The source for obtaining DR spectra, as well as for the Pilot device, was a Xenon arc Lamp. In addition, as in the Pilot device, it was possible to measure spectra for several SDSs: SDS_1 = 400μm, SDS_2 = 600μm, SDS_3 = 800μm and SDS_3 = 1000μm. The SpectroLive instrument was used to measure spectra in a clinical study of non-melanoma skin cancer in humans, the results of which will be described in details in chapter four of this thesis (second main contribution).

1.4. Conclusion

The first chapter discussed the problem of diagnosing non-melanoma skin cancer, the main diagnostic methods that are used in clinical practice, as well as promising alternative diagnostic methods based on the use of optical technologies. The main advantage of choosing optical methods instead of the traditional diagnostic procedure is the possibility of using them for non-invasive diagnosis of skin lesions in real time.

Spectroscopic methods (and some imaging methods) are also highly sensitive to changes in the optical properties of the skin due to carcinogenesis, which makes them promising for diagnosing a wide range of pathologies. At the same time, the high sensitivity of optical methods can be a disadvantage, as the optical properties of the skin vary considerably depending on the individual characteristics of the patient and various influences on the skin throughout his life. All this, together with the noise introduced by the device, makes it impossible to visually diagnose NMSC from spectra and skin images obtained from a large number of different patients.

There are many solutions to the problem of the great diversity in the optical properties of the original skin samples, and as a consequence, of the shapes of spectral curves or image content, but all of them in one way or another lead to the use of different data processing methods. Data processing methods consists of several treatment steps and can produce both positive and negative effects on the interpretability and informativeness of the results. However, a properly selected combination of processing methods and appropriate parameters can allow to significantly increase the accuracy of disease diagnosis. All this makes the processing of experimental data as important as the acquisition of data.

This is especially relevant for optical methods such as AF and DR spectroscopy, as these methods are highly sensitive to changes in skin properties and structure, but also have low interpretability due to the complex structure of the skin and the fact that the fluorescence and absorption peaks of many skin components overlap. Thus, the choice of spectra processing methods needs to be given special attention. As a consequence, the second chapter of the thesis will focus on the use of different data processing methods, to improve the accuracy of NMSC diagnosis. The main focus of the second chapter will be on methods that can be used to process AF and DR spectra, as these methods have been used to increase the performance of the analysis of skin lesions spectra described in chapters three and four.

Chapter 2. Spectroscopic data processing and analysis methods

2.1. Introduction

The first step after obtaining the experimental data is the processing of the spectra to facilitate their analysis. Depending on the purpose of the study, the list of processing methods used may be very different, but in most publications the description of this stage distinguishes two main steps: data preprocessing and data processing.

The output of the data preprocessing stage is a set of data ready for extracting and analysing the information needed for the spectrum analysis. This step is of high importance because many data processing methods are very sensitive to the quality of the input data, as a result, the accuracy of the final analysis of the study can be significantly reduced if the algorithms and approaches for pre-processing the raw data are not properly chosen.

Data processing methods most often include a variety of methods of machine learning, as well as some statistical methods. All Machine Learning (ML) methods can be divided into two main groups - supervised and unsupervised ML methods. In addition, it is customary to distinguish deep learning methods as a specific subcategory of ML methods.

Data processing methods are quite a versatile tool suitable for working with almost any numerical data, which makes their application in spectroscopy highly relevant. However, spectroscopic data have several peculiarities, such as a relatively large number of features (most often from several hundred to several thousand) with a small number of samples, which is especially significant for clinical studies, where researchers often have to deal with the lack of available samples. In addition, the spectra obtained can often be very noisy, which further complicates their analysis and makes the use of data processing techniques, including machine learning methods, necessary.

This chapter will discuss the main steps in the spectroscopic data preprocessing and processing pipeline and the main methods that can be used for each of the steps, including so-called data-fusion approaches that allow more efficient aggregation of results obtained in multimodal studies. In addition, after providing the description of processing of data, it is necessary to define the main accuracy metrics that can be used to assess the performance of its analysis.

2.2. Pre-processing of raw data

The main stages of spectroscopic data preprocessing are: noise removal and filtering, data normalization or scaling, outliers detection and management of features. This last step includes feature extraction and feature selection methods as well as their combinations. Besides, depending on the type of data, other steps can be added. For example, when combining multiple technologies in a single device, it is necessary to make adjustments to the received signal to consider the effects on the signal from different optoelectronic components of the device.

2.2.1. Noise removal

In signal processing, noise is a general term for the unwanted (and, in general, unknown) changes in signals that they may undergo during capture, storage, transmission, processing or conversion [176].

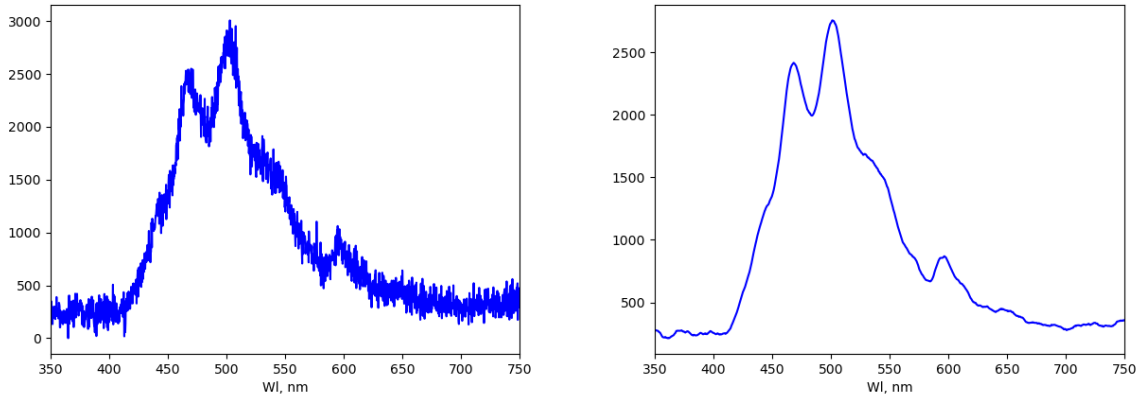
Removing noise is one of the most important stages of data preprocessing, since the use of heavily noisy data can significantly reduce the efficiency of the analysis. In many modern devices, the installed software allows to remove the main types of noise associated with the detector (dark noise) or directly with the signal transmission (read noise). One of the main methods is to increase the exposure time, but this way to proceed also entails a decrease in spectral resolution. Therefore, the exposure time should be chosen based on the wanted Signal to Noise Ratio (SNR) and spectral resolution. Another way to reduce the influence of noise is to subtract a reference signal from the measured resulting signal [177]. These methods often complement each other.

A rather primitive method of removing noise, which is already used in the case of "manual" data processing, is to set a minimum threshold for the data value. All signal below this limit is considered as noise. This approach can be used effectively only in situations where the signal under study is discrete or it is known in advance that the "peak" values of obtained spectra are the most informative, while the rest of the information is much less important. In Raman spectroscopy, it is also necessary to consider the influence of cosmic rays; however, this point will not be considered in this section.

However, it is not always possible to remove all noise by simply subtracting the recorded noise, so filtering methods must be used. In a general sense, filtering is the procedure of removing unwanted components from signals, which are most often understood as time series. There are a large number of different filtering methods and also ways of their

classification: filters can be analogue and digital, infinite impulse response and finite impulse response, passive or active etc.

At the same time, consideration of all listed classifications of filters does not make sense, because in processing of AF and DR spectra the application of filters is limited to two main goals - removal spikes from the obtained spectra and subsequent smoothing. Some examples of filters that can be used for spike removal and smoothing are the median filter and low pass (average) filtering using convolution kernels. In addition to filtering methods, down sampling can also be used for smoothing purpose. An example of spectra before and after removing noise from them is shown in **Figure 2.1**.



(a) (b)

Figure 2.1. Example of spectra before (a) and after (b) removing of noise.

2.2.2. Outlier removing

After removing noise from the data, it is also necessary to check whether outliers are present in the data and, if so, remove them from the set of spectra. This step is also very important, because machine learning methods when training models usually do not use the whole data set, but some part of the samples selected using statistical criteria that underlie the algorithm. The presence of outliers among such samples can have a very negative impact on the efficiency of the subsequent application of the model.

According to [178] there are three main types of outliers:

- Point outliers;
- Collective outliers;
- Contextual outliers;

Point outliers (see **Figure 2.2**) are instances of data that can be considered to be abnormal with respect to the rest of the data. This type of outlier is the most common, and as a result most outlier removal studies and approaches are recognised to deal with point outliers. An example of a point outlier is a spectral curve that has a shape significantly different from all other curves in the dataset.

Collective outliers, may be indistinguishable from the rest of the dataset in terms of statistical criteria, but when considering a set of multiple measurements, it can be inferred that they are outliers. An example of collective outliers is the presence of multiple spectra in a data set with an additional fluorescence peak in a region (e.g., resulting from a failure to follow a protocol for measuring spectra for multiple samples). Examples of point and collective outliers are shown in **Figure 2.2**.

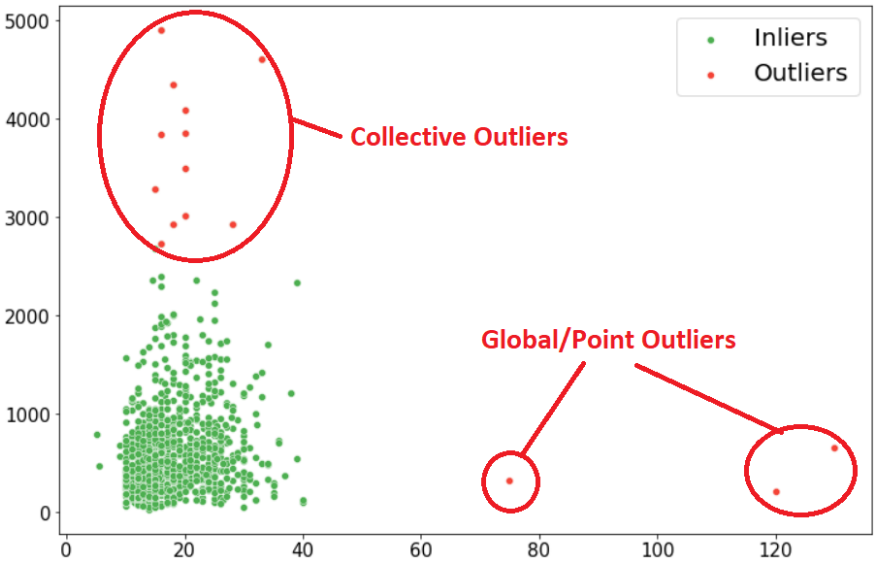


Figure 2.2. Collective and Point outliers [178].

Contextual outliers may be indistinguishable from "the ordinary" data in terms of statistical metrics, but in terms of context, as determined by the structure and properties of the dataset as well as the formulation of the problem to be solved, these instances should be considered as outliers. An example of a contextual outlier would be an air temperature value of 0°C. For cities in eastern France, such a temperature value may be normal in winter, but a similar value obtained in summer for the seasonal dependence of temperature would be a contextual outlier.

Removal of outliers is a necessary step to improve the quality of the analysed results, however, finding out and explaining the causes of appearing outliers is an extremely important

part of the data pipeline, as it can help to interpret the obtained results and to significantly increase the quality of the results obtained in the future.

2.2.3. Normalization, Standardization and Scaling

Data cleaned of noise and outliers may still be unsuitable for analysis by various statistical and machine learning methods, due to the fact that these methods are extremely sensitive to the scale of the data, and may also exhibit significantly reduced efficiency if the data is not optimally represented. To address this problem, Normalisation (which is also referred to as Feature scaling) and Standardisation are commonly used in signal processing.

According to the most common definition, normalization is a method of pre-processing numerical features in the data sets under study in order to bring them to some common scale (usually between 0 and 1) without losing information about the differences in their range [179]. The use of normalization is necessary when analysing the set of data with different units or scales. Standardization, on the other hand, rescales the values to have a mean of 0 and a standard deviation of 1, changing the original value distribution and relative value distances. This is beneficial when comparing data with different means or variances, when applying machine learning algorithms that assume a normal distribution or when implemented algorithms are sensitive to the scale of the data. In this context, the most common normalisation methods are *Standard* Scaling and *MinMax* Scaling:

- The *Standard* Scaler standardizes the data by subtracting the mean value and then scaling the entire signal to the unit variance, which implies dividing all values by the standard deviation.
- For each feature in the set of data, *MinMax* Scaler subtracts the minimum value of the feature and then divides by the range. The range is the difference between the original maximum and the original minimum. The *MinMax* Scaler preserves the shape of the original distribution. It does not change the information in the original data [180].

The methods described above are inherently feature scaling and, if applied to spectra processing, should result in a change in the shape of the spectral curve. However, the definitions of normalisation, standardisation and scaling in spectra processing are slightly blurred, so that dividing a spectrum by its maximum value, wavelength ratio, or, for example, by some reference spectrum, may also refer to normalisation, standardisation and scaling. For example, in [33], Tunnell *et al*. uses standardisation of AF and DR spectra by the ratio of the maximum intensity and the average parameter calculated for an averaged spectrum of healthy skin over

the entire spectra acquisition range (for AF spectra) or in the 630-690 nm range (for DR spectra). An example of the impact of such a standardisation method is shown in **Figure 2.3** [33].

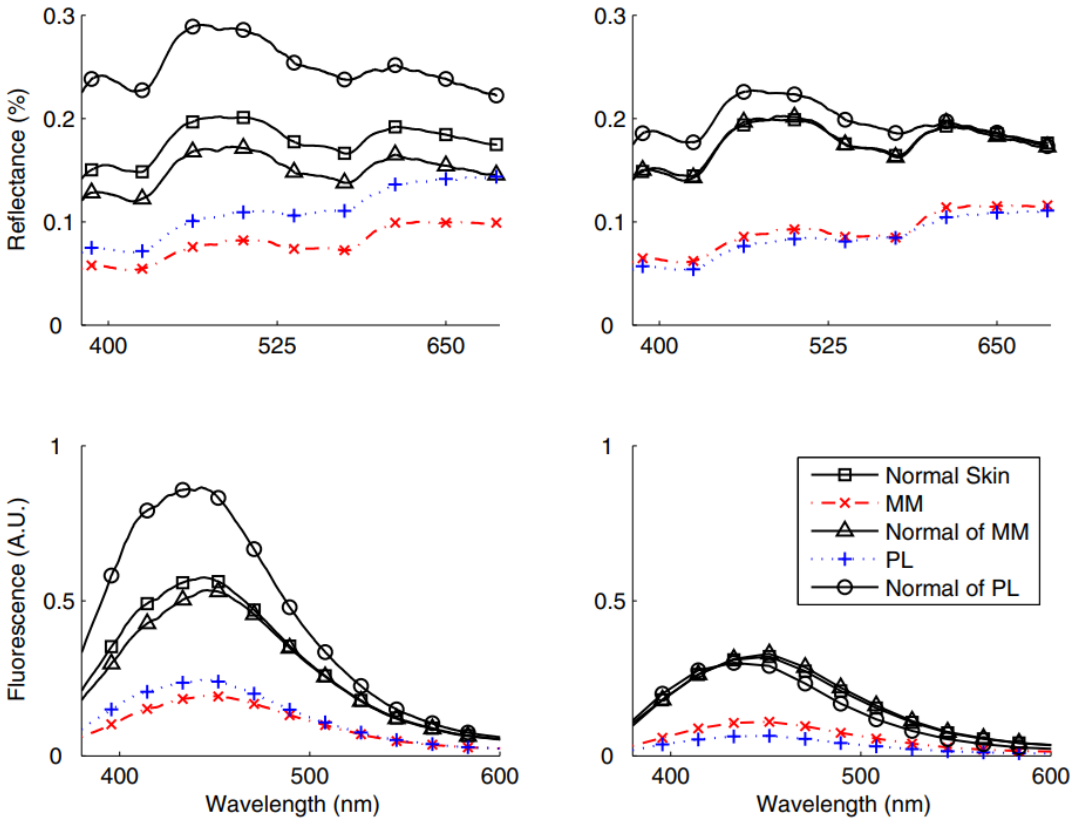


Figure 2.3. Example of normalisation to intensity ratio for AF and DR spectra [33].

The high sensitivity of data analysis methods to the scale of the original data can not only lead to an improvement in the quality of the analysis results after applying standardisation or normalisation, but also to its degradation. As a consequence, the correct choice of normalisation method for each type of data is very important.

2.2.4. Dimensionality reduction

Once the data have been cleaned of noise and scaled, some data analysis techniques become feasible. However, many classification and regression methods perform much worse on data containing a large number of features and relatively small number of samples. In addition, analysing large datasets is time-consuming and computationally intensive.

In statistics and machine learning, features are usually understood as any measurable properties of samples involved in the analysis (e.g., intensity for different wavelengths in the spectrum). The problem of effective reduction of the number of features is solved with the help

of dimensionality reduction methods, the main function of which is to decrease the number of features with the maximum possible preservation of useful information. Dimensionality reduction methods can be divided into Feature Selection and Feature Extraction methods.

However, instead of using Feature Selection and Feature Extraction methods directly for pre-processed spectroscopic data, “a priori” parameters or features such as total area under curve, peak intensity ratios or peak widths at half-height, the slope of the spectral curve in areas of decreasing or increasing intensity, can be extracted directly from the spectral curves [170].

It's also important to note that when analysing multi-component structures of complex composition, such as human skin, the shapes of the obtained spectral curves can vary greatly depending on individual patient characteristics such as age, phototype, anatomical site and the degree of sun exposure of the measured skin area during the patient's lifetime. All of these parameters have a significant influence on the composition and properties of the skin and, consequently, on the shapes of the spectral curves, so that sensitive spectroscopic techniques such as AF and DR spectroscopy most often require the use of Feature Selection or Feature Extraction methods.

2.2.4.1. Feature selection methods

All Feature selection methods are based on the idea that the original data contain a large number of redundant or irrelevant features that can be removed without much loss of information. Thus, Feature selection methods do not globally change the features themselves, but only remove those that do not carry statistically significant information or do not improve the model.

All Feature Selection methods can be divided into [181–183]:

- Filter methods, which select features based on a performance measure regardless of the data processing algorithm used (e.g. Variance Threshold);
- Wrapper methods, which evaluate a subset of features based on the performance of the processing algorithm (e.g. feature selection based on fitting accuracy);
- Embedded methods, which perform feature selection during the execution of the processing algorithm (e.g. the use of Lasso regularization).
- Hybrid methods, which have been proposed to combine the best properties of filters and wrappers.

The main advantage of feature selection methods over feature extraction methods is preserving data characteristics for interpretability, since extraction of features implies a transition to a new feature space. At the same time, feature selection methods are often characterized by less discriminative power than feature extraction methods.

2.2.4.2. Feature extraction methods

Unlike Feature selection methods, Feature extraction methods do not remove the least significant of the features, but transform the entire dataset into another feature space with the maximum possible preservation of useful information. The most common methods of dimensionality reduction used in the processing of spectroscopic data include algorithms such as principal components analysis (PCA), independent components analysis (ICA), t-distributed Stochastic Neighbor Embedding (t-SNE), and uniform multiverse approximation and projection method (UMAP) [184–186].

In mathematical terms, PCA is an orthogonal linear transformation that transforms data from the original feature space into a new space of lower dimensionality [187]. In this case, the new axes of this space, which are called principal components, represent an orthonormal basis in which different individual dimensions of the data are linearly uncorrelated. At the same time, the first (and each subsequent) axis of the resulting coordinate system is constructed in such a way that the dispersion of data is maximized along it [188]. An example of the realisation of this idea is shown in **Figure 2.4**.

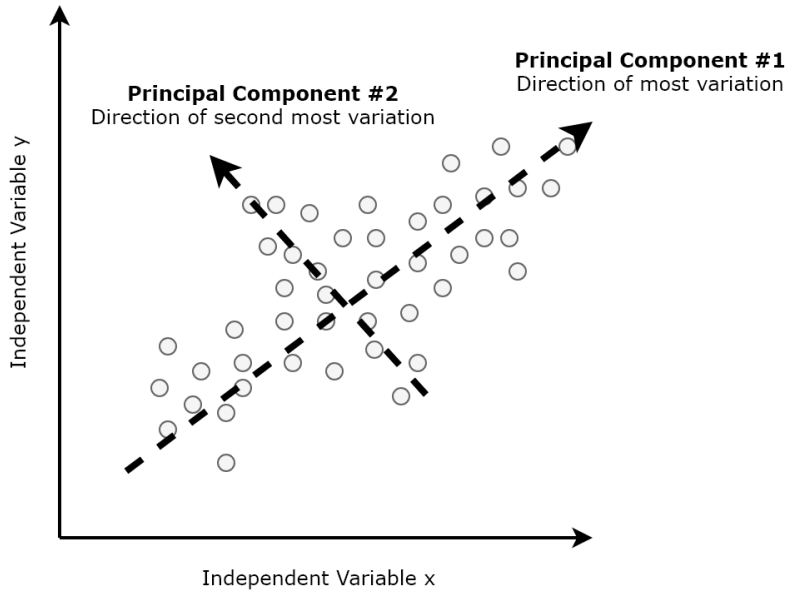


Figure 2.4. Illustration representing two first principal components for a two-dimensional original feature space.

One of the disadvantages of PCA is that the transformation is linear, which does not work well in all cases. This problem can be solved by using a kernel PCA. Kernel PCA uses a kernel function to project dataset into a higher dimensional feature space, where it is linearly separable. It is similar to the idea usually used for Support Vector Machines, which is the most famous method that uses the "kernel trick" [189, 190].

Nonlinear dimensionality reduction can be done in several ways. For example, a low-dimensional surface can be transformed on a high-dimensional space so that a nonlinear relationship among the features can be found. In theory, a lifting function can be used to map the features onto a higher-dimensional space. On a higher space the relationship among the features can be viewed as linear and therefore is easily detected. This is then mapped back on the lower dimensional space and the relationship can be viewed as nonlinear [191]. A close alternative to this approach is the use of kernels, which was mentioned earlier. Another approach to nonlinear dimensionality reduction is the use of manifolds. It is based on the assumption that the data lie on a nested nonlinear manifold, which is smaller in dimension than the original data space, and lies within it. Such methods include Isomap, described in detail in [192].

The t-SNE algorithm consists of two basic steps. At first, t-SNE generates a probability distribution dividing high-dimensional objects into pairs such that similar objects will have a larger probability value, while the probability of selecting non-similar objects will be small. Then t-SNE determines a similar probability distribution over the points already in low-dimensional space, and then minimizes the Kullback-Leibler distance, taking into account the position of the points [193, 194].

UMAP was created to improve the t-SNE algorithm and is very close to it in essence. The principle of this method is to use the concept of k-nearest neighbours and optimize the results using stochastic gradient descent [195].

2.2.5. The use of machine learning methods at pre-processing stage

In addition to the statistical methods described in this section, machine learning methods can also be applied at each stage of data preprocessing. However, when using machine learning methods for noise removal and filtering, it is often necessary to pre-train appropriate machine learning models using pairs of noisy and clean data, which is not always possible since noise free data are not available in numerous applications. When using ML models for outlier

removal, it is sometimes possible to use unsupervised learning approaches such as clustering, but outlier removal based on such a strategy must be explicitly or implicitly justified.

The application of autoencoders for Feature Extraction is quite promising. An Autoencoder (AE) is a simple linear network that is trained without a supervisor and is based on the backpropagation method. The structure of the simplest autoencoder is shown in **Figure 2.5**.

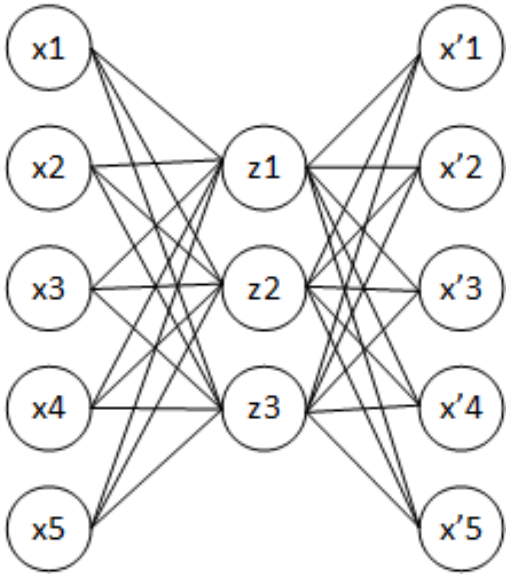


Figure 2.5. Main structure of Autoencoder with one hidden layer.

The main principle of operation and training of the autoencoder is to obtain on the output layer the response as close as possible to the input signal. In this case, for the solution not to be trivial, all inner layers of the autoencoder should be smaller than the outer ones. Thus, by training the autoencoder, it is possible to obtain a transformation of the original data into a new feature space of the same size as the smallest layer of the autoencoder.

2.3. Data processing

The next important step in the pipeline is data processing, which involves transforming the acquired data into meaningful information in accordance with the purpose of the research. The list of possible data processing operations at this stage is quite broad, but the main step is almost always the application of machine learning techniques such as supervised/unsupervised classification or regression. In the context of our work, we are more particularly interested in the AF and DR spectra supervised classification of skin tissues based on the histopathological classification of reference.

2.3.1. Supervised Classification

Machine learning methods used in the frame of optical biopsy studies are most often used to solve the classification problem. Classification is the process of assigning a sample to one of the predetermined sets, called classes. The list of the classes is determined by the application and the task to be solved. Machine learning methods in this case perform the role of a classifier - an algorithm capable of classifying an arbitrary object similar to those used in the training process. Failure to comply with this rule (for example, due to a different set of operations for data preprocessing between the data on which the model was trained and tested) leads to a significant reduction in the effectiveness of the classifier. In addition, like other data processing and analysis methods described above, classifiers are extremely sensitive to the quality of the input data, as well as to the set of hyperparameters under which they operate.

Numerous methods have been used in optical biopsy studies to solve the classification problem, such as Support Vector Machine (SVM) [45], Discriminant Analysis (DA) [122, 196], Logistic Regression (LR) [14, 33, 172], Decision Trees [162, 197] and deep learning methods [13, 17, 157, 198], the most basic of which is the Multilayer Perceptron. It is also worth considering Ensemble classifiers [199], the main idea of which is to synthesise a common prediction based on the use of several simple algorithms.

Support Vector Machine

The main idea of the support vectors machine is to find the dividing hyperplane in the feature space that can maximally separate two classes. A hyperplane is an N-1 dimensional subplane in an N-dimensional Euclidean space, which linearly divides the space into two separate parts [200]. In the case of linearly separable data, once the optimal separating hyperplane is found, the data points lying on its boundary are called reference vector points, and the solution is represented as a linear combination of these points only. Other data points are ignored. Thus, the complexity of the SVM model does not depend on the number of features found in the training data (the number of reference vectors selected by the SVM learning algorithm is usually small). For this reason, SVM is well suited for training problems, where the number of features is large relative to the number of training instances. At the same time, the original SVM is a binary classification method for linearly separable data.

Nonlinear separability of data using SVM is possible by using a “kernel trick”, the essence of which is the implicit mapping of input data into a new multidimensional feature space [201]. The basic idea of the kernel trick is to translate linearly indivisible data from the original feature

space into a new space of higher dimensionality where this data can be linearly separable. Polynomial kernel and Radial Based Function (RBF) are most commonly used instead of linear kernel. An example of using the kernel trick with a polynomial kernel is shown in **Figure 2.6**.

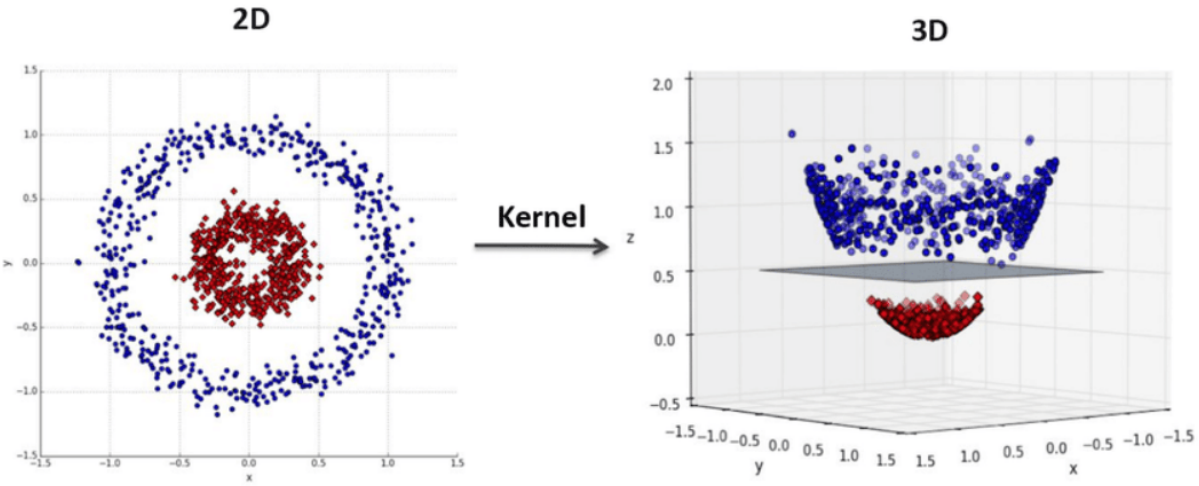


Figure 2.6. Example of using the kernel trick with polynomial kernel.

SVM is originally a binary classifier, which means that in its raw form it can only be used to separate samples into two classes. As a consequence, when using this method for multiclass classification, special solutions must be used. There are two main approaches for performing multi-class classification using SVMs: "one-vs-one" (OVO) classification and "one-vs-rest" (OVR) classification, but these approaches are common to all binary classifiers. The essence of the OVR approach, also sometimes called one-vs-all (OVA), is that each class is compared to a combination of all other classes in turn, and then the results are generalized. In OVO, each other class is opposed to each class, after which the results of all binary classifications are also summarized. Obviously, OVO requires much more computational resources when the number of classes is larger than 3, but in some cases it can work somewhat more accurately. An alternative way to create a multi-class SVM by combining binary classifiers is the Binary Decision Tree [202], which is to build a hierarchy of multiple two-class SVM classifiers as shown in **Figure 2.7**. It is also important to note that similar strategies can be applied to other binary classifiers.

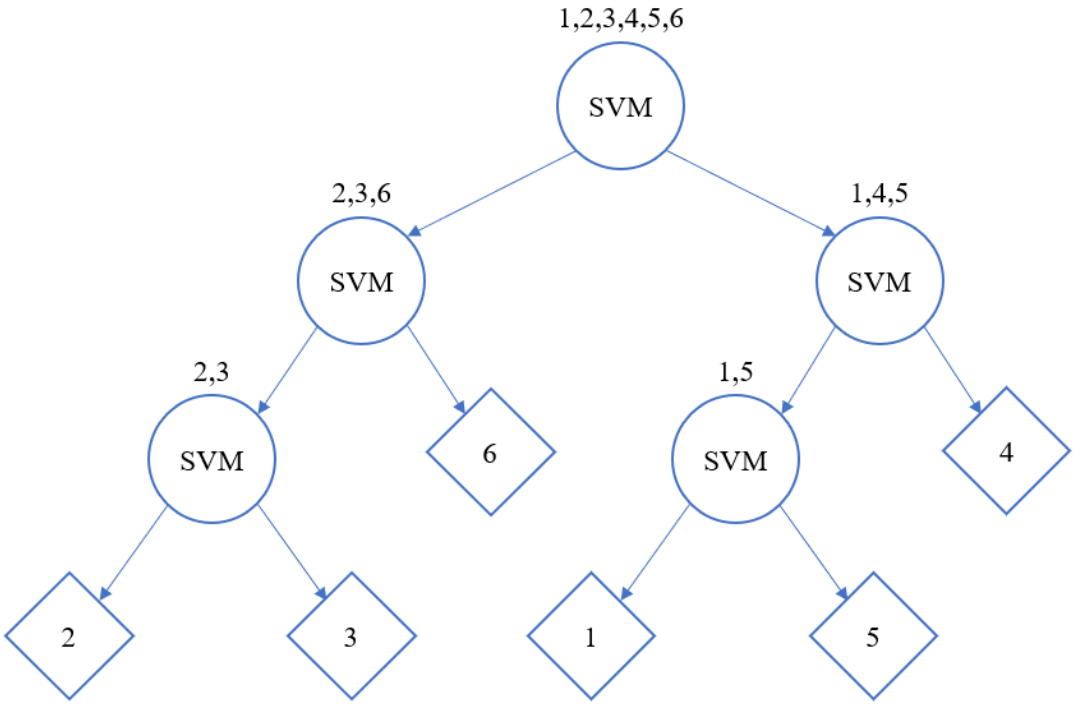


Figure 2.7. Example of Binary Decision Tree implementation for a dataset with six classes.

It is also important to note that multiclass implementations of SVM also exist [203–206]. Their implementation shows similar results compared to methods combining multiple two-class classifiers, and in some cases even outperforms them [207, 208]. However, the implementation of a multi-class SVM requires more time and computational cost than the implementation of a two-class method.

In addition, it should be noted that SVM in any variations is very sensitive to the scale of the data, which requires a careful choice of the pre-processing methods.

Discriminant analysis

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant analysis [209]. The main idea of LDA is to find a linear combination of features that characterizes or separates two or more classes of data. The resulting model can be used as a classifier or, more commonly, as a dimensionality reduction method before subsequent classification. LDA is closely related to PCA as both methods are looking for linear combinations of variables that best explain the difference between data points in the initial dataset. The principal difference between the methods is that LDA explicitly tries to model the difference between classes of data, while PCA, in contrast, does not account for any class differences.

A closely related classifier to the LDA is the quadratic discriminant analysis (QDA). Both LDA and QDA are based on Bayes' rule [210]. If the QDA model assumes that the covariance matrices (matrices composed of the pairwise covariances of the elements of one or two random vectors) are diagonal, then it is assumed that the input data are conditionally independent in each class, and the resulting classifier is equivalent to a Gaussian naive Bayesian classifier. LDA, on the other hand, is a special case of QDA, for which it is assumed that all classes have the same covariance matrix.

The advantages of the method are the simplicity of implementation and interpretation of the results, as well as the low sensitivity to scaling. The main disadvantage is that for complex data with a large number of mutually dependent features, this classifier may not be applicable.

It is also important to note that LDA, like SVM, is a binary classifier, so when using LDA for multiclass classification tasks, it is also necessary to apply OVO or OVR techniques.

Logistic regression

Another example of a binary classifier (in its original form as well), is logistic regression (LR), which is a statistical model that models the probability of one event out of two alternatives [211, 212]. In this case, the logarithm of the probability of a particular outcome for an event is a linear combination of one or more independent variables. The process of learning in this case implies the selection of coefficients at given variables, which consists in maximizing the likelihood function (maximum likelihood method). The name of the method comes from the logistic function that transforms logarithmic coefficients into probability. Most often, this method also includes the use of regularization (see **Section 2.2.2**) to avoid overfitting the model. For multiclass classification, the OVR approach is used.

LR is one of the simplest and best known supervised machine learning algorithms, however, this method is not always suitable for analyzing large datasets without explicit separability.

The main advantage of the method is the comparative ease of implementation and interpretation of results. It’s disadvantages include high sensitivity to the type of feature distribution.

Decision Trees

The decision tree (DT) [213] is a node-based hierarchical structure, where at each node the classified data set is divided according to some decision rule [214]. Accordingly, all the

“branches” coming out of each node represent a test result, and each final node represents a “leaf” associated with a class label.

The main advantage of decision trees is simplicity for understanding and interpreting the results. Moreover, DTs can be combined with other classification methods. However, DTs often have relatively low accuracy and low stability, because with a small change in the dataset used to train the classification model, the decision tree structure can change significantly. In addition, single trees often exhibit a tendency to overtraining A solution to this problem is possible by using ensemble methods, the most common of which is Random Forest (RF).

A random forest combines multiple decision trees into a single classifier, with each tree containing a randomization step consisting of two parts. The first step of randomization is that, the data set applied to each tree is different each time, and samples may be repeated. The second is that a different set of features, ranging in size from the maximum number of features to the full size of the whole dataset, may be applied to each node. Thus, RF is a DT-based ensemble classifier.

Combining results from different trees is achieved either by voting, or by averaging the results over all trees. This approach is a special case of “bagging” (shortened term standing for “Bootstrap” aggregating). Therefore, the variance for each individual tree is reduced, which involves a decrease in the classification accuracy individually, but for the whole forest, the classification accuracy is higher. The disadvantages of the random forest compared to a single decision tree are the greater difficulty in interpreting the results and longer calculation time.

Multilayer perceptron

A MultiLayer Perceptron (MLP) classifier is a fully connected artificial neural network of direct connection (ANN), which means that the nodes of the neural network do not form a loop [215, 216]. Unlike a conventional perceptron, this classifier is able to separate objects that are not linearly separable. A multi-layer neural network consists of large number of units (neurons) joined together in a pattern of connections. Units in a network are usually segregated into three classes: input units, which receive information to be processed; output units, where the results of the processing are found; and units in between known as hidden units.

During classification, the signal on the input units is distributed throughout the network, determining the activation values on all output units. Each input unit has an activation value, which represents some feature external to the network. Each input unit then sends its activation value to each of the hidden units to which it is connected. Each of these hidden units calculates

its own activation value and this signal is then sent to the output units. The activation value for each receiving unit is calculated according to a simple activation function.

Neural networks are trained by processing examples, each containing known "inputs" and "outputs," forming probability-weighted associations between them, which are stored in the data structure of the network itself [215]. Training a neural network is usually done by determining the difference between the processed network output (often a prediction) and the target output, which is the value of error. The network then adjusts its weighted associations according to the learning rule and using that error value. Consistent corrections will cause the neural network to produce an output that is increasingly similar to the target output.

The main advantage of this type of classifiers is that it is capable of processing even large datasets without explicit separability, while obtaining high accuracy. However, properly determining the size of the hidden layers is a challenging task, because an underestimate of the number of neurons can lead to poor approximation and generalization capabilities, while excessive nodes can result in overfitting and eventually make the search for the global optimum more difficult. Also, with a large network, the computation time is significantly higher than for SVM. In addition, this method, like SVM, is very sensitive to data scaling.

The multilayer perceptron is a basic classification model belonging to the set of deep learning approaches. Deep Learning methods refers to as separate category of machine learning methods based on the use of neural networks of different architectures with a large number of hidden layers. The use of deep learning shows good results in many domains such as computer vision, segmentation or new data generation. However, optimization of deep neural networks is computationally and time consuming, and their performance on small one-dimensional datasets is comparable to traditional machine learning algorithms.

Ensemble classifiers

Ensemble methods combine several base models, such as decision trees or logistic regression, in order to obtain more accurate prediction than that obtained by using each single model individually. Simple algorithms in this case are called "weak learners" and the ensemble method is called "strong learner". The most common ensemble methods are based on voting, stacking, bagging, and boosting [199, 217].

There are two main types of voting: “hard” voting, in which class labels for samples are predicted using majority rule voting, and “soft” voting, in which class labels are based on

maximum values of sums of predicted probabilities [218]. The “soft” voting approach is recommended for an ensemble of well-trained classification models.

Stacking involves adding another classifier (meta-learner) that works based on predictions from weak learners [219]. The main advantage of stacking and voting is the possibility to use different types of classifiers as weak learners, and the disadvantages are still more complicated interpretation of the results. However this problem is typical for all ensemble methods.

The basic idea of boosting is to iteratively teach many of the same weak learners to create a strong learner. Boosting is similar to bagging in this aspect. The difference is that each successive model learns on the basis of the previous one, trying to improve its result [199]. There two of the most common algorithms for boosting include adaptive boosting and gradient boosting. Adaptive boosting uses an iterative method to calculate the optimal weights in the model. The algorithm first trains the first base model (say decision trees) on a training set. The relative weight of incorrectly predicted values is increased. The updated weights are fed to the input of the second model and the model is trained, then the predictions are produced and the cycle repeats. Gradient boosting differs from adaptive boosting in the way that weights are optimized, because gradient descent is used in this case .

Ensemble methods are not necessarily capable of significantly improving results relative to the best weakest learner, which is especially noticeable for small datasets (up to a few hundred samples).

2.3.2. Regression

Regression methods must be used when, based on the input data, the values of one or more continuous variables have to be predicted instead of only the value of one discrete variable (class). The solution of such regression task is to establish a relationship between the dependent variable and set of independent variables and to predict new values of the dependent variable based on this relationship. Regression analysis in the case of only one independent variable is very close to the curve fitting process, but it focuses more on statistical inference issues [220].

There are many regression methods, each with its own advantages and disadvantages for solving particular tasks. The most commonly used methods include Linear Regression, Polynomial Regression, Lasso Regression, Ridge Regression, and Elastic Net Regression [221–225].

Linear regression

Linear regression is the simplest regression method and can be described using equation (2.1):

$$Y = \sum_{i=1}^{n} a_i \cdot X_i + b + \varepsilon \quad (2.1)$$

where Y is an dependent variable, X_i are independent variables, a_i are weights of independent variables, b is intercept, n is a number of independent variables and ε is an error. The coefficients of the regression model are fitted with the aim to minimize ε. Root mean square error is usually used as a metric of error minimization. Linear regression works reasonably well when the relationship between the dependent and independent variables is linear, the values of the dependent variable do not contain a large amount of error, and the variance is constant and relatively small.

Polynomial regression

For cases when the relationship between the dependent and independent variables cannot be described as linear, it is reasonable to use polynomial regression [223]. The main difference with linear regression is that in the non-linear regression the degrees of the multiplicative coefficients of the independent variables can exceed 1:

$$Y = a_1 \cdot X_1 + a_2^2 \cdot X_2 + \cdots + a_n^k \cdot X_n + b + \varepsilon \quad (2.2)$$

where Y, X_i, a_i, b, n and ε have the same meanings as in equation (1), and k is the degree of the coefficient on the independent variable X_n. As in the case of Linear regression, the model is trained by minimising the mean square error. Polynomial regression can model more complex non-linear relationships between variables, however, if the maximum degree of the polynomial is incorrectly chosen, such a model may be more prone to overfitting.

Ridge regression

Linear and polynomial regressions demonstrate high efficiency when the set of independent variables is non-collinear, which means that the independent variables do not have linear dependencies among themselves. However, in some tasks, the independent variables can often be collinear, which can lead to overfitting and thus reducing the accuracy of the model. In this case, the method of training a regression model based on minimizing the mean square error cannot be used. The problem of training a model on data sets containing collinear features can be solved using several methods, one of which being Ridge Regression [224].

The main idea behind ridge regression is to add a penalty to the sum of squared errors to be minimized, based on the weights of the independent variables such as:

$$Arg_min = \left([Y - a \cdot X]^2 + \lambda \cdot \|a\|^2\right) \quad (2.3)$$

where Y is a matrix of values of the dependent variable, X is a matrix of values of the independent variable, a weight coefficients on independent variables and λ is a regularization coefficient. This approach introduces a penalty for too large coefficients, making the algorithm train on all variables more uniformly, reducing thus the impact of collinearity of the independent variables on the final model. By changing the value of λ it is possible to adjust the value of the penalty for the values of weights in order to optimise the performance of the method.

In general, the method of adding any constraints to the problem solving condition (in this case, to the regression model training condition) is called regularisation. More specifically, this approach used in Ridge Regression is called Tikhonov regularisation or *L2* regularisation. It is also important to note that in the case of *L2* regularisation, the absolute weights for the independent variables can be as small as possible, but will never reach 0, which mean that there will be no selection of features. An alternative method is *L1*-regularisation, which isalso implemented in Lasso regression [224].

Lasso regression

Like Ridge Regression, Lasso Regression introduces a penalty for too large values of the weights of the independent variables, thus preventing overfitting (Equation (2.4)).

$$Arg_min = ([Y - a \cdot X]^2 + \lambda \cdot \|a\|) \quad (2.4)$$

All variables are similar to those presented in Equation 3. However, unlike the Ridge Regression, the weights on the independent variables in the penalty term are taken to the first degree rather than squared. As a consequence, for the same value of λ, Lasso Regression imposes a smaller penalty for the same values of the weights than Ridge Regression. Another important difference from Ridge Regression is that the absolute values of weights for independent variables can be equal to 0, which means that Lasso Regression also performs feature selection in the process of model training, having a positive effect on the interpretability of the resulting model.

Elastic net regression

As highlighted in the previous paragraphs, Ridge regression is best applied when it is necessary to consider the influence of all independent variables on the final result, whereas to Lasso regression is applied when it is necessary to take into account the influence of only a few independent variables, and the remaining variables can be discarded [225, 226]. However, in some tasks it is not known in advance which regularisation method performs best. In such cases, it is reasonable to use Elastic net regression, which combines both approaches (see Equation (2.5)).

$$Arg_min = \left([Y - a \cdot X]^2 + \lambda_1 \cdot \|a\| + \lambda_2 \cdot \|a\|^2\right) \quad (2.5)$$

It can be seen that instead of imposing a single penalty on the values of weight coefficients at independent variables, the minimized value for Elastic net regression imposes two penalties at once, corresponding to L1 ($\|a\|$) and L2 ($\|a\|^2$) regularizations. By changing the values of coefficients λ_1 and λ_2 it is possible to modify the impact of each type of regularization, thus optimizing the resulting model appropriateness.

2.3.3. Unsupervised Classification (Clustering)

Unsupervised Classification (Clustering) is the task of grouping a set of objects in such a way that objects belonging to one group (cluster) are more similar to each other than to objects belonging to other clusters [227]. There is no precise definition of what exactly is a cluster, as it strongly depends on the algorithm. Popular notions of clusters include groups with small distances between cluster members, dense areas of the data space, intervals or particular statistical distributions. The main types of clustering methods include [228]:

- Clustering methods based on connectivity, also known as hierarchical clustering, which are based on the idea that objects are more related to nearby objects than to objects further away (e.g, single-linkage clustering);
- Centroid-based clustering methods, when each cluster is represented by a central vector, which is not necessarily a part of the dataset (e.g, k-means);
- Clustering methods based on probability distribution. Clusters in this case are defined as objects belonging most likely to the same distribution;
- Density-based clustering methods, when clusters are defined as areas of higher density of samples than the rest of the dataset, such as Density-Based Spatial

Clustering of Applications with Noise (DBSCAN) or Ordering Points To Identify the Clustering Structure (OPTICS).

An important stage of clustering lies in the to choice of a method with its appropriate parameters so that, on the one hand, the method makes it possible to divide the data into a sufficient number of clusters and, on the other hand, does not create new data clusters where they should not be according to the problem conditions. In addition, the clustering results should be consistent with the available information about the dataset.

For diagnostic purposes, clustering can be used to confirm the separability of data belonging to different diagnostic classes or subclasses, as well as to find new classes and subclasses in the data according to clinical or diagnostic information [229]. At the same time, since it's more appropriate in most diagnostic studies to address the supervised classification problem, clustering methods will not be discussed in detail in this thesis.

2.4. Data fusion

In general, all tasks that demand any type of parameter estimation from multiple sources can benefit from the use of data fusion methods. According to [230] data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences than that could be achieved by the use of a single sensor alone.

There are a large number of different Data Fusion methods and ways to categorize them [231]. However, in our context of combining multimodal spectroscopic data more efficiently, it is convenient to distinguish three main categories based on the stage of the data pipeline at which the fusion is performed:

- Fusion of the input data;
- Fusion of features;
- Fusion of decisions.

When using Fusion of the input data or Fusion of features, some of the common features for different modalities are removed or lose their importance, which may negatively affect the accuracy of subsequent analysis.

An alternative approach is decision fusion, where in the first step each modality is analyzed independently, after which the data analysis results are combined by some method (e.g. majority voting or stacking). This approach reduces the risk of losing features for

individual modalities. However, accounting for the relationship between modalities in this case is quite difficult. In addition, the risk of overtraining increases.

2.5. Quantitative evaluation of the performance of data analysis algorithms

In order to select the optimal methods of data processing and analysis, it is necessary to define metrics that can be used to evaluate the effectiveness of each combination of methods and corresponding hyperparameters for classification and regression tasks.

2.5.1. Statistical metrics for classification

Most of the statistical metrics that can be used to evaluate the performance of a classification task will be explained in the case of a binary classification. Binary classification is a task where all samples from a dataset need to be divided into two classes. For diagnostic tests their labels correspond to a ‘positive’ and ‘negative’. All the most commonly used statistical metrics are based on using all possible predictions for each sample: number of true positives (TP), number of true negatives (TN), number of false positives (FP) and number of false negatives (FN). All accuracy metrics in binary classification tasks, which will be presented next, are calculated using TP, TN, FP and FN values. The most convenient and informative way to present the results in this way is to use the confusion matrix, which is shown in **Figure 2.8**.

		Predicted class	
		Positive	**Negative**
Actual class	**Positive**	TP	FN
	Negative	FP	TN

Figure 2.8. Example of a confusion matrix.

False positive and false negative outcomes are also referred to as errors of the first and second kind, respectively. An error of the first kind is often referred to as a false alarm. This term is commonly used in medical diagnosis of diseases and means that the model has predicted a disease where it is not actually present. With a large number of false positives, the collateral damage of such a system can outweigh the benefits. At the same time, errors of the first kind are less dangerous in medical practice than errors of the second kind (false negatives), because

in the case of the second kind of error, a patient with a disease is recognised by the system as healthy, which may lead to untimely diagnosis of the disease and, as a consequence, to a significant complication of treatment or even death of the patient.

All the accuracy metrics discussed below apply to the case of binary classification, as this greatly simplifies their description. At the same time, almost all similar metrics are also applicable for multiclass classification by averaging each metric over all classes, or calculating them for each class separately.

Accuracy

The most simple and straightforward metric for assessing the performance of a classification model is the accuracy quality criterion, which shows the proportion of correctly classified samples out of the total number of samples. The accuracy can be calculated using equation (2.6).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.6)$$

The advantages of the accuracy are simplicity and versatility, which makes it possible to use it to evaluate the effectiveness of almost any classification models. At the same time, when using only accuracy, it is impossible to take into account the false alarm ratios of the model, which can be very critical in some cases. For example, when choosing between two models for diagnosing a disease, it is possible to have a situation in which the accuracy of the first model is slightly higher, but the second model has a much lower proportion of false negative predictions, making the second model preferable for application. Another important disadvantage of accuracy is that it does not always work correctly with unbalanced data sets.

Sensitivity and Specificity

A possible alternative or rather an addition to the use of accuracy or a possible complement to it could be sensitivity and specificity, which are calculated using equations (2.7) and (2.8).

$$Sensitivity = \frac{TP}{TP+FN} \quad (2.7)$$

$$Specificity = \frac{TN}{FP+TN} \quad (2.8)$$

If we consider the application of sensitivity and specificity on the example of medical diagnosis of a disease, then, as it can be seen from the formulas, that sensitivity corresponds to the proportion of samples with a disease for which the diagnosis was correctly determined.

Specificity shows the proportion of healthy samples that were correctly identified as healthy by the model.

Sensitivity and specificity provide much more information about the quality of the model, however, they should only be used together, as the fact that sensitivity alone or specificity alone are high enough does not mean that the model as a whole provides satisfactory results.

Precision and Recall

Another way to assess the quality of the model is to use a combination of Precision and Recall, which can be found using equations (2.9) and (2.10).

$$Precision = \frac{TP}{TP+FP} \tag{2.9}$$

$$Recall = \frac{TP}{TP+FN} \tag{2.10}$$

As can be seen from the equations, Recall is calculated in the same way as sensitivity, and Precision characterizes the proportion of results correctly predicted as 'positive' from all samples predicted as "positive". This approach may be somewhat less informative than the previous approach, as it does not provide direct information about what proportion of the 'negative' class samples were correctly predicted as "negative". Instead, the Precision and Recall combination provides information on what the chance is that the 'positive' diagnosis for an unlabelled sample is correct.

Receiver Operating Characteristic curves

Another way to evaluate the quality of a classification model is to use ROC (Receiver Operating Characteristic) curves, which graphically represent the relation between True Positive Rate (TPR) and False Positive Rate (FPR) values, that are calculated using equations (2.11) and (2.12).

$$TPR = \frac{TP}{TP+FN} \tag{2.12}$$

$$FPR = \frac{FP}{FP+TN} \tag{2.13}$$

where TPR corresponds to Sensitivity and Recall, while FPR shows the proportion of "negatives" misclassified as "positives".

In case of binary classification, the algorithm has to select a threshold of the class variable values above which all data will belong to the "positive" class and below which all

data will belong to the ‘negative’ class (or vice versa). The graph called ROC-curve can be obtained by manually changing the threshold of the class variable from -∞ to +∞ and plotting the values X = FPR and Y = TPR in the X,Y coordinate space. In addition, when the values of the class variable threshold are equal to -∞ and +∞, the model classifies all samples only as “positive” or only as “negative” respectively. Thus, the ROC curve always passes through the points (0,0) and (1,1). An example of the ROC curve is shown in **Figure 2.9**.

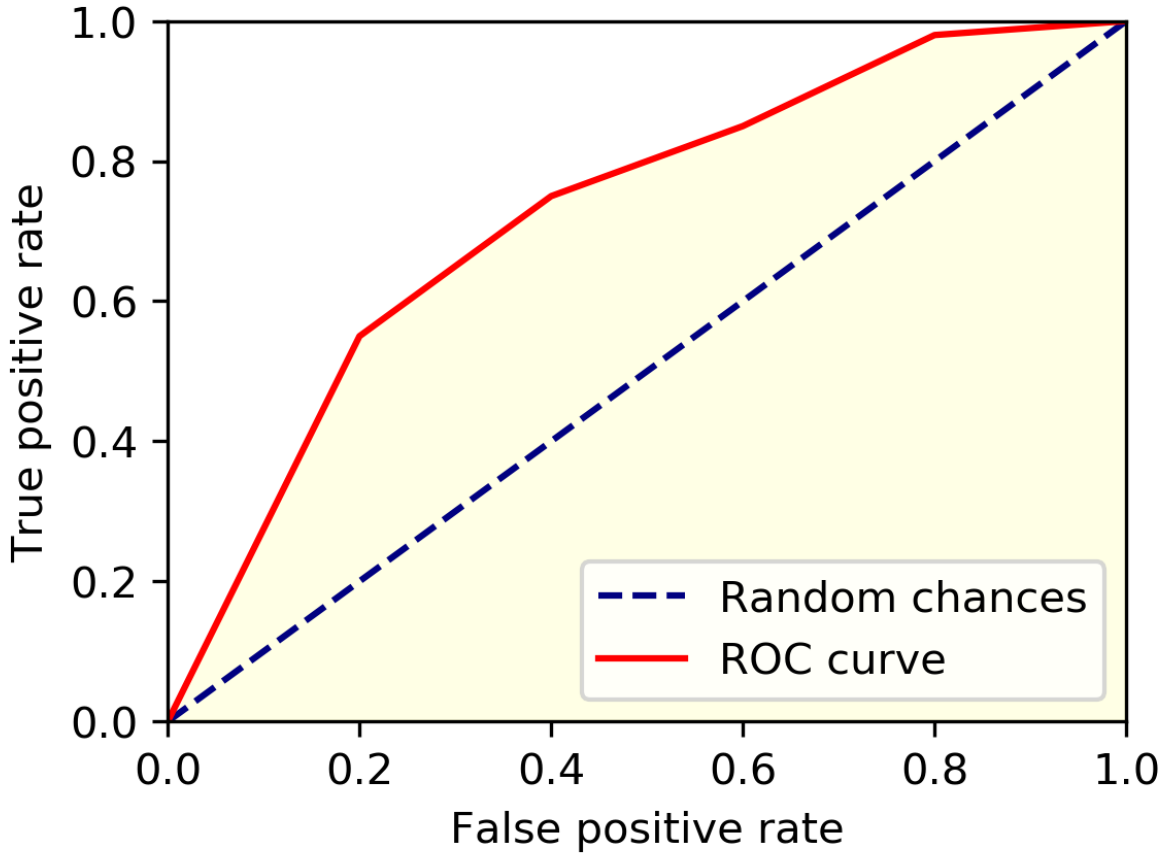


Figure 2.9. Example of a ROC-curve.

The ROC curve resulting from a binary classification should always lie above the straight line segment connecting points (0,0) and (1,1), as it corresponds to random guessing. If the curve is close to or coincides with the segment line of random guessing, it means that the classification algorithm in its current configuration is not suitable for this task. Theoretically, if the ROC curve lies well below the random guessing line, it means that the classifier for some reason acts in the opposite way and if the class names are swapped, the model will perform better. However, in practice, algorithms for training classification models should not lead to this situation.

The ROC curve can be used to find the point with the optimal balance between TPR and FPR and thus determine the best value of the class separation threshold for a class variable. Comparison of different ROC curves is usually done using the area under the curve (AUC), which shows the probability that a randomly selected instance of a negative class is less likely to be recognized as a positive class than a randomly selected positive class.

Other metrics

In addition to using ROC curves, there are other methods of combining several simple parameters for estimating model performance into one, such as the Youden's index (or Informedness) and the F_1-score, which can be found using equations (2.13) and (2.14).

$$\text{Youden's index} = Sensitivity + Specificity - 1 = \frac{TP}{TP+FN} + \frac{TN}{FP+TN} - 1 \quad (2.13)$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (2.14)$$

The use of the Youden index reduces the number of parameters for evaluating the effectiveness of the classification model. However, this parameter does not distinguish between false positive and false negative predictions, so it is usually not used in the tasks of medical diagnosis of various pathologies.

The F_1 score is the harmonic mean of the precision and recall. It thus symmetrically represents both precision and recall in one metric. However, this parameter does not consider the difference in importance between Precision and Recall. When it is necessary for the purposes of the study, F_β score is used instead of F_1 score, which can be calculated using equation (2.15).

$$F_\beta = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (2.15)$$

The value of the parameter β varies depending on which of the parameters is more important for solving a particular problem.

2.5.2. Accuracy metrics for regression

The most commonly used accuracy metrics for regression tasks are related to the estimation of the mean fitting error, because unlike classification problems, in regression, the output or projected features are continuous. All accuracy metrics for regression problems presented below are based on differences between y_i - theoretical data for the i-th sample, $\hat{y}_i$ - predicted data for the *i-th* sample, as well as average data values for the whole dataset [232, 233].

Mean Absolut Error

The Mean Absolute Error (MAE) is the average of the absolute difference between the observed value and the predicted values [233, 234].

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |y_i - \hat{y}| \tag{2.16}$$

where n is the number of predictions

The main advantage of MAE is its simplicity. However, in cases where the data set contains both positive and negative numbers, MAE can reduce the actual magnitude of the error. For this reason, quadratic errors are used in most studies.

Mean Squared Error and Root Mean Squared Error

The Mean Squared Error (MSE) is determined with the squared differences between the actual values and the predicted values and RMSE is the square root of MSE [233].

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^{n} |y_i - \hat{y}|^2 \tag{2.17}$$

The main advantage of RMSE is that it is measured in the same values as the original data.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} |y_i - \hat{y}|^2} \tag{2.18}$$

R-Squared (R^2)

R-squared is the proportion of the variance in the dependent variable that is explained by the independent variables. It can be calculated using equation (2.19) [233, 235].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{2.19}$$

2.5.3. Splitting of the data set

In addition to the proper choice of accuracy metrics, it is necessary that the sets of data that are used to train and test the model do not contain the same samples in order to correctly evaluate the performance of the resulting classification and regression models. For this purpose, the original dataset is usually divided into two or three parts:

- Training set that is used to train the model parameters;
- (optional) Validation set, which is used for model validation during the model type selection and hyperparameter optimization phase;

- Test Set that is used to validate the model and is not involved in the learning process at all;

This split of the data set is optimal, however, in most optical biopsy studies the data set size is between a few tens to a few hundreds of samples, which may often be not enough to train valid machine learning models. In this case, a validation set is usually not used, and as a consequence, the entire model optimisation process is based on the prediction results for the test set.

The test dataset is almost always smaller than the training dataset (except for the case of working with very large datasets, the training of models based on which requires huge time and computational resources). The ratio of training and test set sizes is usually in the range from 1 to 4, and increasing this ratio increases the risk of overtraining the model.

It is important to note that the choice of which samples belong to the training set and which other samples belong to the test set should be done randomly (taking into account additional factors if necessary), and the whole sequence of operations from the moment of splitting the samples into sets to the moment of validation of machine learning models should be repeated several times, as randomly chosen splitting of samples into training and test sets can provide both highly overestimated and highly underestimated results. This approach basically illustrates the idea of cross-validation.

Cross-Validation (CV) is a repeated validation method that uses different parts of the dataset to test and train the model in different iterations [236]. Two types of cross-validation can be distinguished: exhaustive and non-exhaustive cross-validation.

Exhaustive cross-validation methods are cross-validation methods that examine and test all possible ways of dividing the original sample into a training and validation set. Types of exhaustive cross-validation methods include "leave-*p*-out" cross-validation (LPOCV) and "leave-one-out" cross-validation (LOOCV). Leave-*p*-out cross-validation involves using *p* observations as the validation set and the remaining observations as the training set [237]. This is repeated on all ways to cut the original sample on a validation set of *p* observations and a training set.

Leave-one is a particular case of leave-p-out cross-validation with $p = 1$. The main disadvantage of this type of methods is the large computation time. Non-exhaustive cross validation methods do not compute all ways of splitting the original sample. Those methods are approximations of leave-p-out cross-validation.

In *k*-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k - 1$ subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data [238]. The k results can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling (see below) is that all observations are used for both training and validation, and each observation is used for validation exactly once. *5-fold* and *10-fold* cross-validation are commonly used, but in general k remains an unfixed parameter. When $k = n$ (the number of observations), k-fold cross-validation is equivalent to LOOCV. In stratified k-fold cross-validation, the partitions are selected so that the mean response value is approximately equal in all the partitions. In the case of binary classification, this means that each partition contains roughly the same proportions of the two types of class labels. The general scheme of a k-fold (with $k = 5$) CV is presented in **Figure 2.10**.

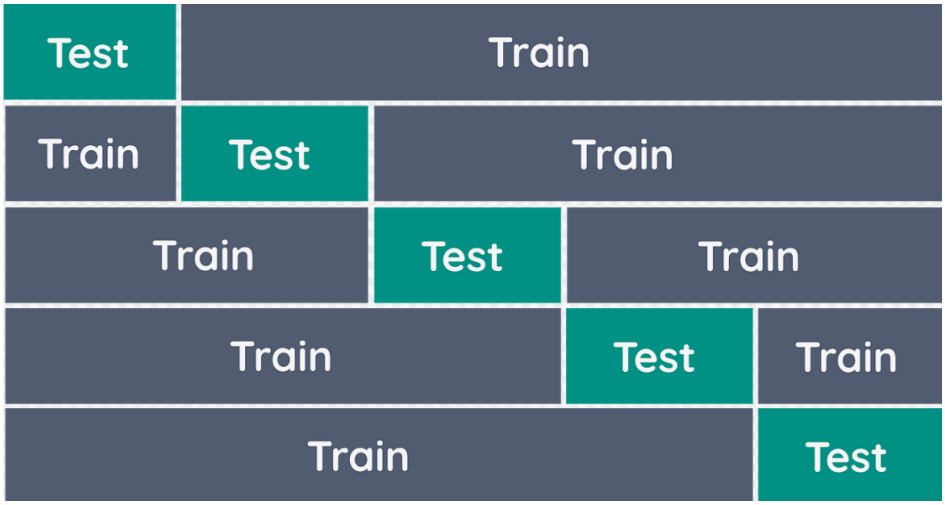


Figure 2.10. Example of *k*-fold cross validation for $k = 5$.

Repeated random sub-sampling validation, also known as Monte Carlo cross-validation, creates multiple random splits of the dataset into training and validation data [239]. For each such split, the model is fit to the training data, and predictive accuracy is assessed using the validation data. The results are then averaged over the splits. The advantage of this method (over *k*-fold cross validation) is that the proportion of the training/validation split is not dependent on the number of iterations (i.e., the number of partitions). It's drawback is that some observations may never be selected in the validation subsample, whereas others may be selected more than once. In other words, validation subsets may overlap. This method also exhibits

Monte Carlo variation, meaning that the results will vary if the analysis is repeated with different random splits. As the number of random splits approaches infinity, the result of repeated random sub-sampling validation tends towards that of leave-p-out cross-validation.

When cross-validation is used simultaneously for selection of the best set of hyperparameters and for error estimation (and assessment of generalization capacity), a nested cross-validation is required. At the same time, the basic principles of functioning of this type of cross-validation are identical to those described before.

2.6. Conclusion

The second chapter introduced the problem of the spectroscopic data processing for diagnostic studies. The need to apply data processing and analysis methods, including machine learning methods, to spectroscopic data is caused by the complex structure of biological tissues and low interpretability of different types of raw spectra, including AF and DR spectra.

The list of existing methods of data processing and their functions is very wide, and for almost every method there is a list of customized hyperparameters. As a consequence, the proper selection of the list of used data processing and analysis methods, as well as the configuration of appropriate hyperparameters for each method, is an important stage of any study.

When processing and analysing multimodal data, an additional important aspect is to make the best possible use of all available modalities in obtaining the final result of the study. The methods used to combine data from different sources for the same samples are called Data Fusion, and can be applied at different stages of the data pipeline. However, as with other Data Processing and Analysis methods, the proper selection of Fusion methods and their optimization has a significant impact on the final result.

Another important step is the selection of accuracy metrics used to compare the performance of different combinations of methods and parameters for classification or regression tasks. The chosen accuracy metrics should, on the one hand, provide sufficient information about the performance of the model and, on the other hand, be understandable and interpretable.

After reviewing the basic steps of spectroscopic data processing, it becomes possible to validate these methods on the data of a pilot study in which the possibility of classifying benign and precancerous skin lesions in mice using spatially resolved DR and multi-excited AF spectroscopy was investigated.

Chapter 3. Pre-clinical study on mice skin

3.1. Introduction

As it is clear from the previous chapters, combining optical biopsy and data processing techniques is a rather promising direction in NMSC diagnosis. However, the implementation of projects in this field is always fraught with certain difficulties, primarily related to the need to perform a large number of trials, including preclinical trials on animals or *ex vivo* tissues and clinical trials on humans.

Preclinical studies are an extremely important aspect aimed at providing a preliminary practical justification of the effectiveness of the chosen experimental methodology. At this stage, the basic strategy for processing and analysing the data obtained is also usually developed and optimised. Clinical tests of the developed methodology on humans, in turn, are necessary for the final confirmation of its effectiveness in the conditions in which the diagnostic method will be used.

Following all formal procedures as well as recruiting the required number of patients takes a large amount of time (up to decades). In addition, based on the results of clinical and preclinical tests, it may be necessary to make any adjustments to the experimental device parameters or data processing pipeline, which may also increase the already lengthy process of implementing a diagnostic method.

Since the ultimate goal of this thesis was to develop a methodology for analysing spectra measured *in vivo* from human skin, the main idea of the preclinical study was to obtain spectra measured on animal skin to test and optimise the spectra processing pipeline. At the same time, as previously stated, conducting such a study from the ground up could have taken years. However, the CRAN team had already conducted a preclinical study on mouse skin using a bimodal spectroscopic device. Thus, it was decided to use the spectra obtained in this study to test and optimise the data analysis pipeline.

This chapter will describe the materials and methods of the preclinical study as well as the results obtained. Particular attention will be paid to comparing the performance of different combinations of dimensionality reduction, classification and data fusion techniques.

3.2. Materials and methods

3.2.1. Characteristics of the experimental device

The pilot device was used in the series of preclinical studies to take spectra from the skin of mice with precancerous and benign skin lesions [42, 43, 46, 47]. The results of the last preclinical study [47] will be presented in chapter three of this thesis. The main scheme of the device is shown on **Figure 3.1** [47].

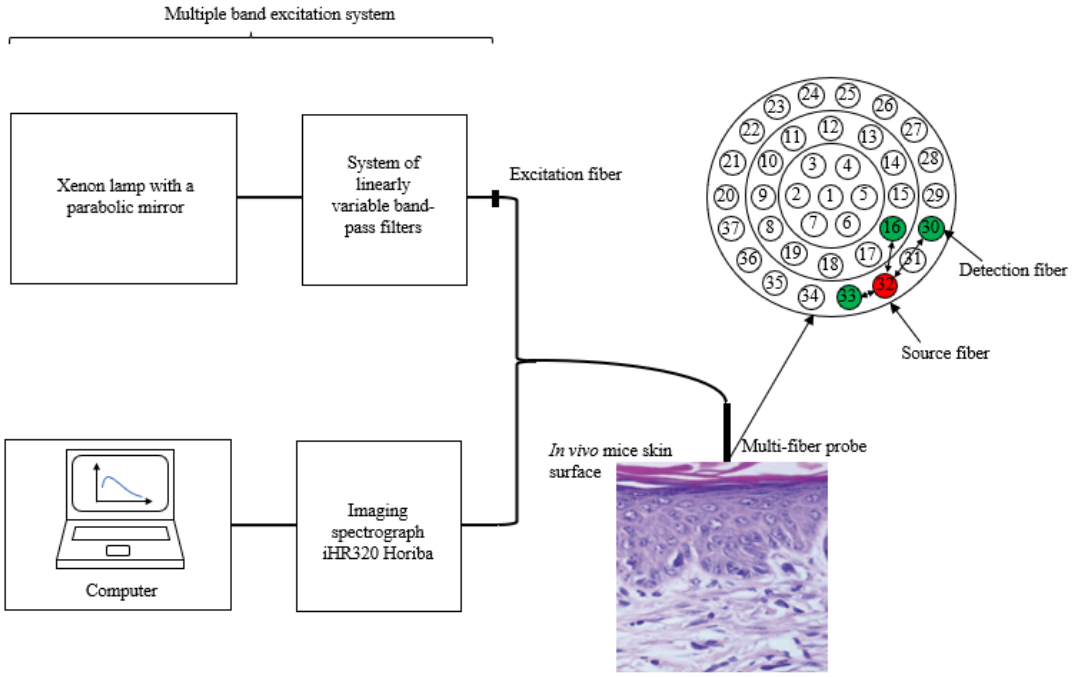


Figure 3.1. Scheme of the pilot device used in preclinical study.

Xenon lamp emitting mainly in the 300–800 nm spectral range and a system of linearly variable band-pass filters were used to obtain DR and multi-excitation AF spectra corresponding to different excitation wavelengths. The light was focused into the source fibre core, that contains 37 optical fibres (numerical aperture is 0.22, SEDI, France) arranged in concentric circles within the 2 mm-diameter bundle. For this application, three different source-detector separations were selected: 271μm (SDS_1), 536μm (SDS_2) and 834μm (SDS_3) for distances between the excitation fibre #32 and the detection fibres #33, #16 and #30, respectively (see fig. 3.1). Increasing the SDS allows you to increase the proportion of photons recorded from the lower layers of the skin, which can be a significant advantage in diagnostic tasks. However, with increasing SDS, the total number of detected photons decreases, as a result of which the signal-to-noise ratio becomes smaller, which makes it difficult to analyse the obtained spectra. The use of multiple SDS in this device is a compromise between the ability to obtain more information in the spectra and complicating their analysis.

The latter were selected among the bundle fibres as close as possible to the probe border in order to be more easily located by the investigator and precisely positioned onto the skin measurement area targeted. Main characteristics of the device are shown in **Table 3.1**, including AF excitation peak wavelengths and average output (incident) power used during the study.

Table 3.1. Main light source and detection characteristics of the device used during the experimental study (*FWHM = Full Width at Half Maximum).

Modality	**Source wavelength (nm)**	**Output power (µW)**	**Spectra acquisition range (nm)**	**Source-detector separations (µm)**
Diffuse reflectance spectroscopy	Ranges: 370–540 450-640 560-740	 430 305 242	390-720	SDS_1 = 271 SDS_2 = 536 SDS_3 = 834
Multi-excitation Autofluorescence spectroscopy	Peak (FWHM*): AF_1 = 360 (17) AF_2 = 368 (17) AF_3 = 390 (17) AF_4 = 400 (15) AF_5 = 410 (15) AF_6 = 420 (15) AF_7 = 430 (15)	 19 25 32 30 30 26 20		

The spectra collected at three SDS were simultaneously measured by an imaging spectrograph. Each measure was repeated three times and then averaged at the end of pre-processing stage to improve the signal-to-noise ratio. These averaged spectra were used as the initial data for further processing.

3.2.2. Characteristics of the dataset under study

The experimental protocol of this study was approved by the French Ethical Committee on Animal Experimentation. Twelve-week-old female hairless mice were divided into two groups: the control group with 8 sham-irradiated mice and the test group included 20 mice, that were irradiated by UV with a fluence of 3 mW/cm^2 for 20 seconds each time from the dorsal side [43]. Spectra were collected on the backs of mice at several constant points to increase the chance of reaching the lesions.

The data set consisted of AF and DR spectra measured *in vivo* on N = 246 mice skin samples. Based on the results of the histopathological analysis, all mouse skin samples were classified into four classes: healthy skin (H), compensatory hyperplasia (CH), atypical hyperplasia (AH) and dysplasia (D).

It should be noted that atypical hyperplasia presents the risk of evolving into dysplasia, which in turn can develop into benign neoplasia or squamous cell carcinoma. Whereas compensatory hyperplasia is not a precancerous state of skin. Therefore, classifying all the spectra as potentially precancerous (PP) and not at risk is also of some interest. The characteristics of the data set under study are presented in **Table 3.2**.

Table 3.2. Main characteristics of the initial set of spectra obtained from mice skin.

Class	Label	Prognosis	Number of samples
Healthy	H	Not at risk	84
Compensatory hyperplasia	CH	Not at risk	47
Atypically hyperplasia	AH	Potentially precancerous	59
Dysplasia	D	Potentially precancerous	56
		Total	**246**

Compensatory hyperplasia is characterised by the proliferation of keratinocytes, so mitoses can be detected in the upper layers, whereas "normally" they are limited to the basal layer. Hyperproliferation is accompanied by strong metabolic activity of keratinocytes, which leads to increased keratin synthesis, as a result of which the keratin layer covering the epidermis increases significantly.

Atypical hyperplasia corresponds to the "intermediate" state of the tissue, i.e. the stage when abnormal morphological elements become visually distinguishable, but no accumulations were detected. Among the abnormal effects, chromatin heterogeneity in the nuclei was mainly distinguished, with the upper layers still containing nuclei, which indicates cytological dismaturation and disorganization of cell layers.

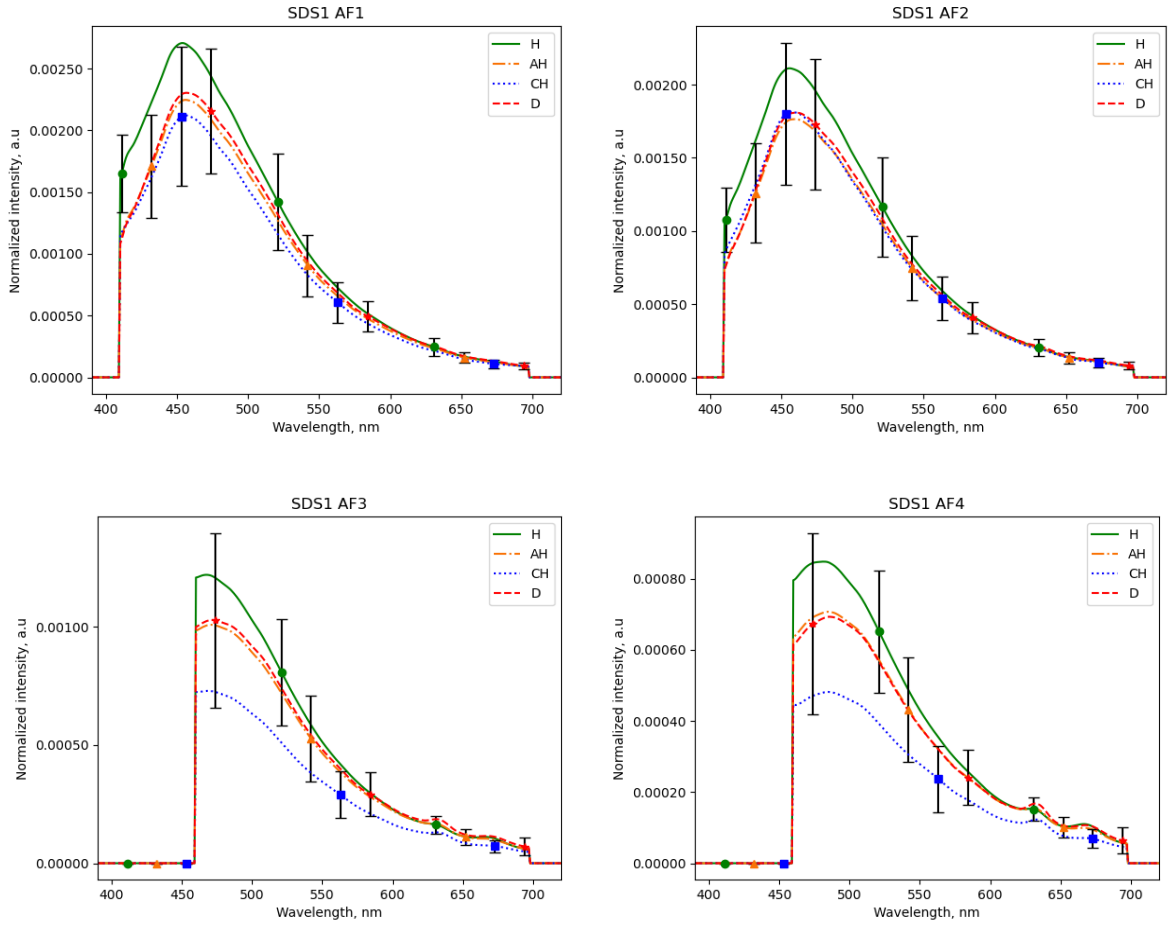
In case of Dysplastic hyperplasia (or "Dysplasia") in addition to the presence of atypical cells mentioned before, there were characteristic features of keratinocyte dysfunction (parakeratosis and dyskeratosis) and dermal fibroblasts (elastosis). Parakeratosis corresponds to a keratin layer that still contains keratinocytes or cell nuclei, whereas normally the stratum corneum consists only of fully mature keratin filaments. Dyskeratoses correspond to fully mature (differentiated) keratinocytes in the lower layers, whereas normally keratinocytes fully mature only on the surface, in the stratum corneum. Similarly, fibroblasts of the dermis damaged by UV radiation synthesize abnormal elastic fibres: short and thick. Thus, atypical hyperplasia and dysplasia have similar symptoms and have similar effects mostly on the morphology of the tissue and, as a consequence, on its scattering properties. However, the effect of dysplasia on skin properties is stronger.

The main symptoms of the skin states under study, as well as their effect on skin properties, are presented in **Table 3.3**.

Table 3.3. Symptoms of the skin states under study and their effects on mice skin properties.

Diagnosis	Main symptoms	Cause	Layer of skin
CH	Hyperproliferation of keratinocytes	Increased epidermis thickness	Epidermis
AH	Cytological dysmaturation, disorganisation of cell layers	Local changes in cell morphology	Epidermis
D	Disorganisation of cell layers, Dyskeratosis, Parakeratosis, Elastosis	Keratinization, extensive change of a layer organization in the epidermis	Epidermis, dermis (elastosis only)

A description of the diagnostic criteria for all classes and some examples of microscopic images of the mice skin with different types of hyperplasia, were presented in [43]. The mean spectra corresponding to all classes obtained for all possible combinations of modality and SDS are presented in **Figures 3.2-3.4**. In addition, **Figures 3.2-3.4** shows the standard deviations obtained by averaging the spectra by class.



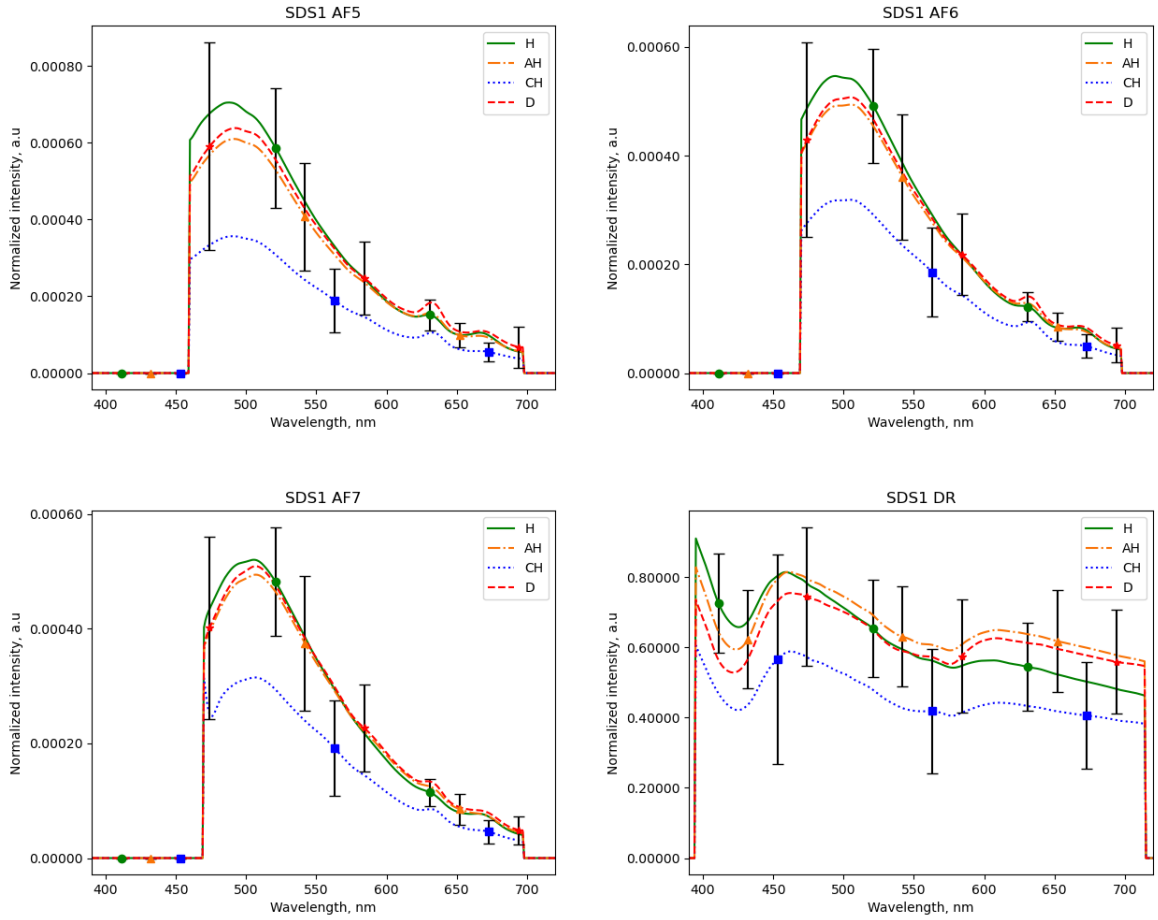
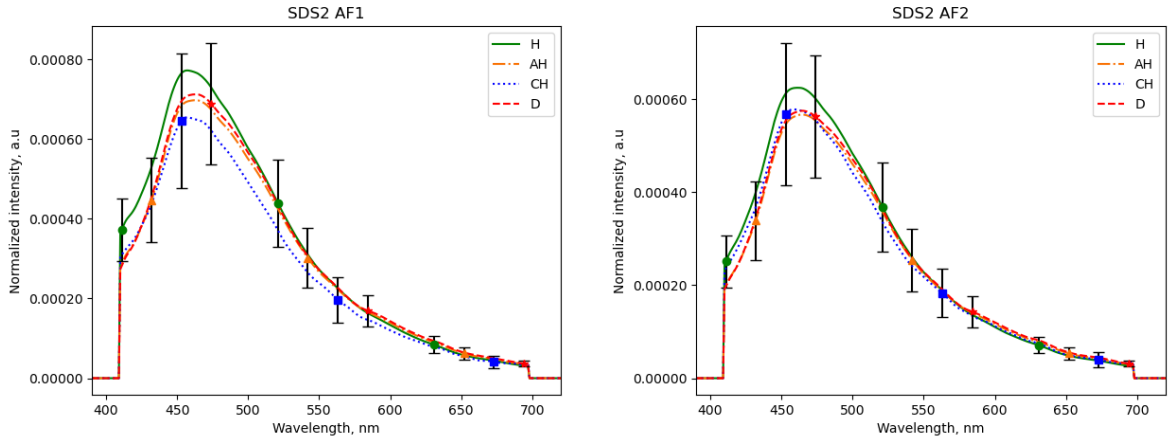


Figure 3.2. Mean spectra for all classes obtained at SDS_1 = 271μm.



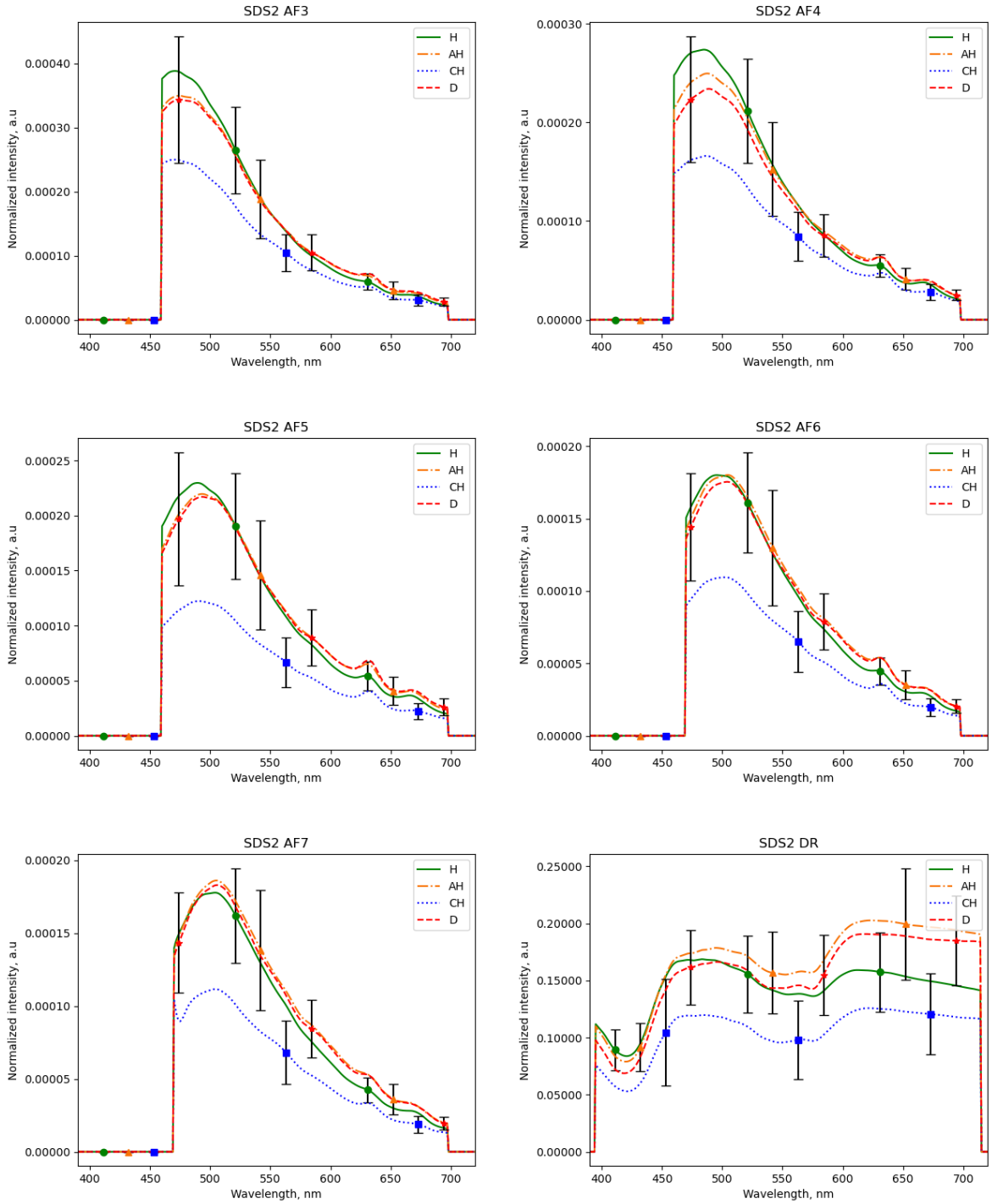
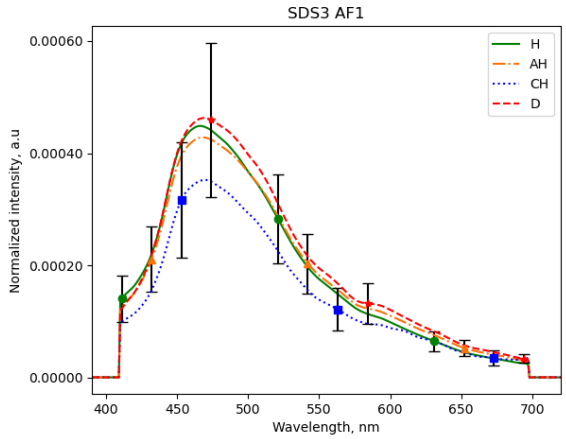
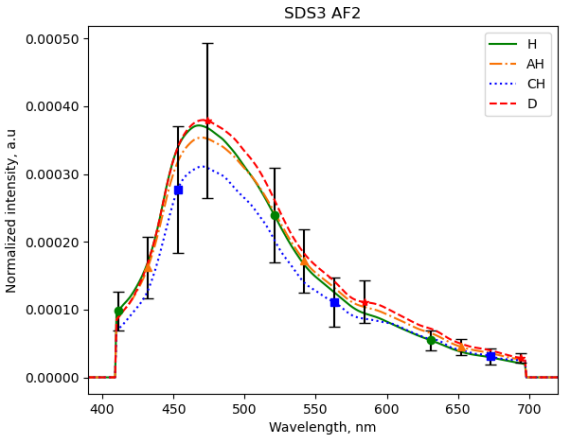
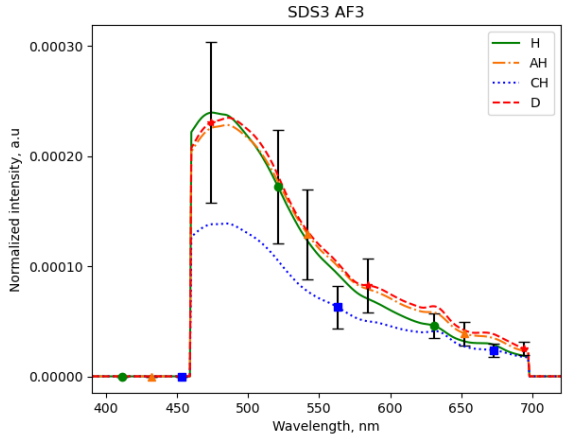
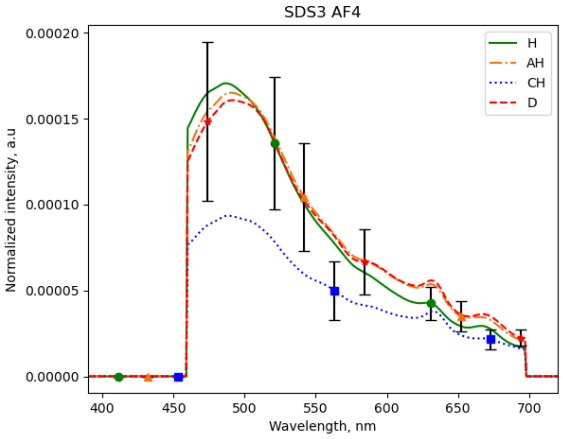


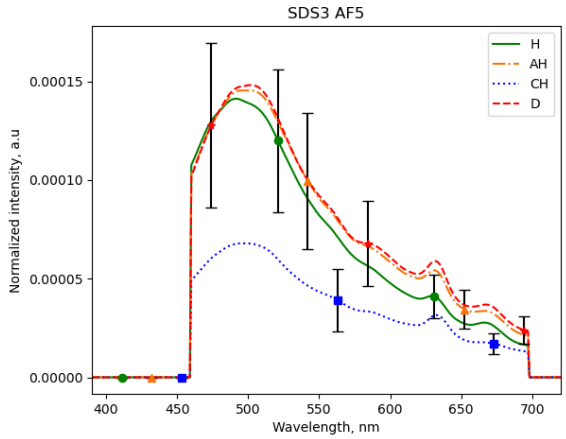
Figure 3.3. Mean spectra for all classes obtained at SDS_2 = 536μm.

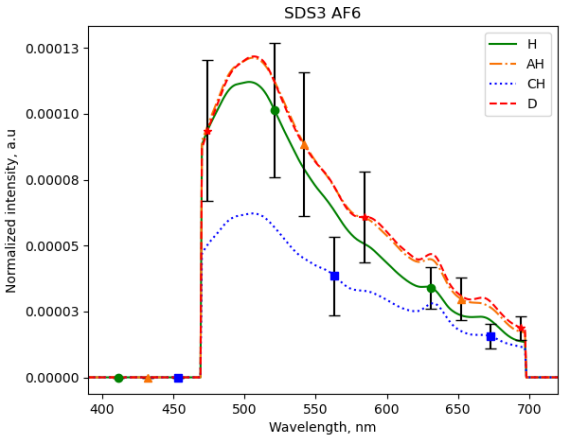

SDS3 AF1
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D


SDS3 AF2
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D


SDS3 AF3
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D


SDS3 AF4
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D


SDS3 AF5
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D


SDS3 AF6
Normalized intensity, a.u
Wavelength, nm
H
AH
CH
D

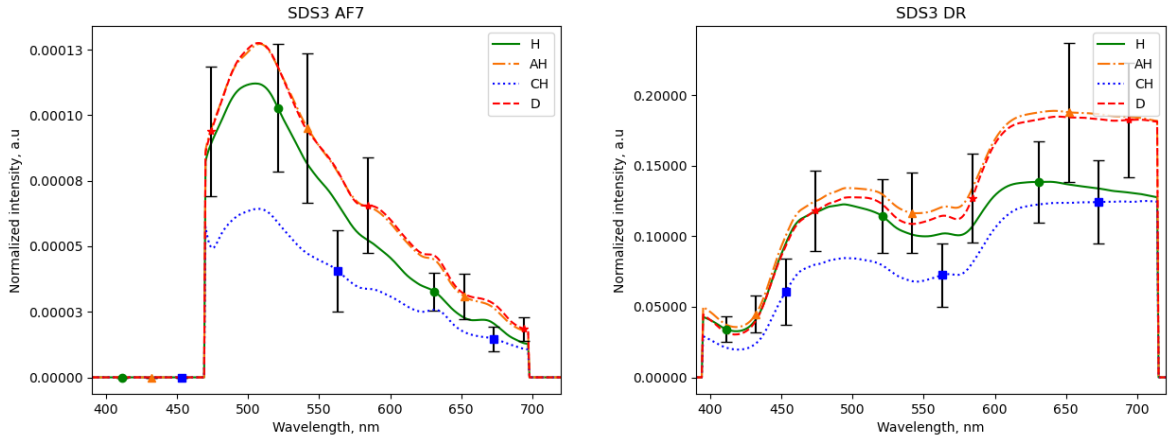


Figure 3.4. Mean spectra for all classes obtained at SDS_3 = 834μm.

It can be seen that for almost all of these classes the areas of standard deviations overlap significantly, making the classification of the spectra a rather difficult task. This problem can be dealt with by implementing data processing techniques in combination with machine learning methods for classification.

3.2.3. Spectrum analysis pipeline

The general scheme of the spectroscopic data analysis pipeline includes pre-processing, classification and data fusion steps, due to the use of two modalities (DR and AF with seven excitation wavelengths), and three SDS values, which resulted in 24 groups of spectra under study: one diffuse reflectance spectra and seven autofluorescence spectra with different excitation wavelengths for each of all three SDS.

3.2.3.1. Pre-processing of the spectra

The initial spectra pre-processing stage consists of three steps: noise removing, dimensionality reduction/feature extraction for every group of spectra and normalization.

At the stage of noise removal from the obtained spectra, a median filter was used to remove noise peaks and a Savitzky-Golay filter was used to smooth the spectral curves. After application of filters for additional smoothing, downsampling from a step of 0.3 nm to 1 nm was also used. After that, for each series of three spectra measured at the same point for the same source and SDS, averaging was performed to improve the signal-to-noise ratio.

The principal component analysis was used for feature extraction, however, unlike the previous studies, the PCA was applied to each of the initial spectra rather than to a set of manually extracted spectral features. The number of principal components in each case ranged from 2 to 50. It is important to note that the new features derived from PCA are uncorrelated,

which is particularly important for further classification step. In addition, several kernels were tested for PCA: linear, polynomial and radial basis function (RBF).

Standard Scaler and *MinMax* Scaler functions were proposed as normalisation methods. It's also should be noted that the final result of the classification can be influenced not only by the method of normalisation, but also by the stage at which it is implemented. To find the best combination of normalization method and stage of its application, all the normalization methods were applied before and after the application of PCA. The final scheme of spectra pre-processing is shown in **Figure 3.5**.

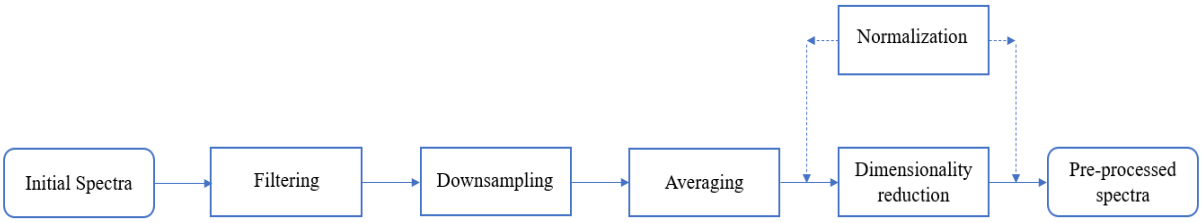


Figure 3.5. The main scheme of spectra pre-processing.

3.2.3.2. Classification of the spectra

After dimensionality reduction step, classification of the spectra with the optimization of hyperparameters for each classifier was carried out using SVM, MLP, LDA and RF. All classifiers were tested in parallel in order to select the most accurate one and corresponding set of hyperparameters. For SVM, which is a binary classifier, different approaches have been also tested for multi-class classification: one-vs-one (OVO) classification when each class is opposed to each class, after which the results of all binary classifications are also summarized and one-vs-rest (OVR) when each class is compared to a combination of all other classes in turn, and then the results are generalized. A list of hyperparameters is given in the **Table 3.4**.

Table 3.4. List of classification methods and corresponding hyperparameters used in preclinical study.

Name of the method	List of optimized hyperparameters	Variation of the hyperparameters
Support vector machine (SVM)	Kernel	Linear/Polynomial(2,3)/RBF
	Regularization parameter value	0.1-10
	Decision function type	OVO/OVR
Multi-layered perceptron (MLP)	Number of hidden layers	1-4
	Size of each hidden layer	10-100
	Type of the activation function	Identity/logistic/rectified linear unit (ReLU)
Linear discriminant analysis (LDA)	Type of solver	SVD/LSQR

A stratified k-fold cross validation (CV) was used to evaluate the classification performance. The k-value was varied from 3 up to 20 in order to find the optimal value. In addition, leave-one-out cross validation (LOOCV) was performed when SVM and LDA were used for classification.

The evaluation of the effectiveness of classification was based on two quality criteria: the multiclass classification accuracy (equation 3.1) and the error in classifying precancerous lesions as healthy (equation 3.2):

$$Acc = \frac{TP_{CH}+TP_{AH}+TP_{D}+TP_{H}}{N_T} \times 100\% \qquad (3.1)$$

where TP_{i} stands for correctly classified results for class i and N_T refers to as the total number of samples.

$$Err = \frac{FP_{AH-H}+FP_{AH-CH}+FP_{D-H}+FP_{D-CH}}{N_{AH}+N_D} \times 100\% \qquad (3.2)$$

where $FP_{\mathrm{i\text{-}j}}$ represents misclassified results corresponding to class i incorrectly predicted as class j and N_i is the number of samples for class i.

3.2.3.3. Data Fusion

Once classification was performed for all groups of spectra obtained for each unique combination of source and SDS, data fusion was performed in several ways. The first approach

was to concatenate linearly the spectra obtained at each point for all combinations of the source configuration and SDS before performing feature extraction and normalisation.

The next approach was based on combining features obtained after PCA for each group of spectra, and was applied after the use of PCA. The effectiveness of normalisation was tested both before and after the use of this type of data fusion. Thus, the subsequent classification was based on features extracted from the spectra obtained for all sources and SDSs together, rather than individually.

The last group of fusion approaches was carried out by analysing the results of the classification of every of 24 groups of spectra. This approach is also called decision fusion. The most standard method used for decision fusion is majority voting. However, when using majority voting it is not possible to take into account differences in the accuracy of the original predictions, which may have a negative impact on the final result. To avoid this problem, two other methods were tested to combine the decisions in parallel with majority voting: stacking and Weighted Voting (WV).

Another important point is that the different accuracy of the predictions for all 24 groups of spectra, as well as the correlation between them, leads to the fact that using all the data at once can also reduce the final classification accuracy. For this reason, the data were combined in three ways: combining all groups of spectra at once (1 result), by modality/AF excitation wavelength of the spectra (8 results) and by their SDS (3 results). In addition, for the last two cases, decision fusion was repeated again for the new predictions obtained.

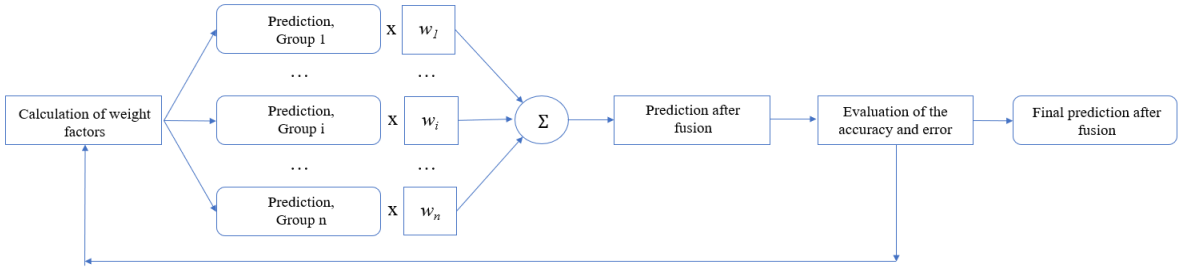


Figure 3.6. The main scheme of Data Analysis by using Weighted Voting.

The general scheme for WV (**Figure 3.6**) includes calculation of the weights (w_i) for each selected group of spectra, following by the aggregation of the contributions of all groups to the final diagnosis for each sample (Σ). The weights were changed for all groups and their combinations from 0 to 1 in steps of 0.1 after normalisation of their values to make the sum of the weights equal to 1. After summing up the contributions from each group of spectra, a

prediction was generated and the accuracy of the prediction was then evaluated. If the classification accuracy was lower than the best value obtained earlier (or the error value was higher when the accuracy was equal), the values of the weights were changed. The process was stopped after going through all the values of weights.

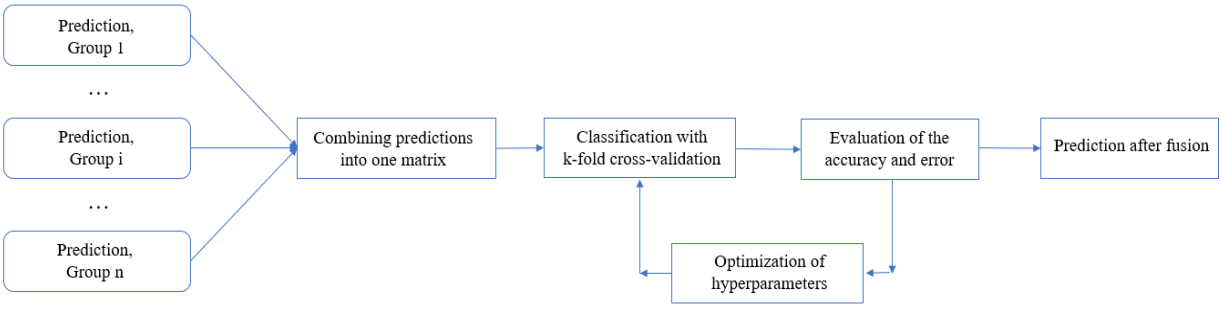


Figure 3.7. The main scheme of Data Analysis by using Stacking.

The general scheme of stacking which involves adding one more classifier that is used do analyse decisions obtained by the first step of classification as new set of features, is shown on **Figure 3.7**. The classification and optimization strategy is similar to the one described in **Paragraph 3.2.2**.

3.3. Results and Discussion

3.3.1. Results of methods and hyperparameters selection

The process of selection the hyper-parameters was carried out based on the accuracy of the classification. The first step involved the selection of an optimal number of principal components used in the study, as well as the selection of the kernel for PCA, that provide the best accuracy of classification.

An example of a 3D distribution of the thirst three principal components for DR spectra (SDS = 536μm) for all the classes is shown on **Figure 3.8**.

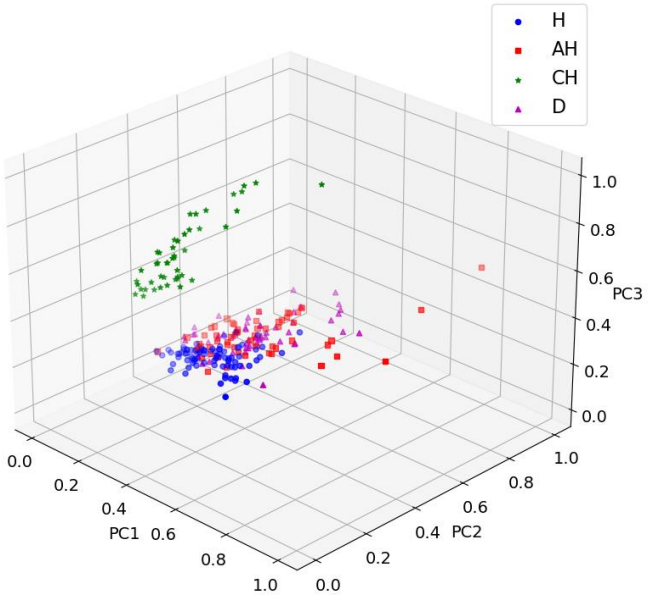


Figure 3.8. 3D distribution of the first three principal components for all four classes of spectra.

It can be seen that the point clouds corresponding to each class are overlapped. This result is repeated for all other modalities/sub-modalities and SDSs, consequently, no visual separation between all classes could be achieved. The use of other combinations of principal components also did not provide any effect, therefore it was decided to use a larger number of principal components for the classification. An example of the dependency between the accuracy of multi-class classification using SVMs and the number of principal components (from 2 to 80) for DR spectra obtained at SDS_2 is shown on **Figure 3.9**.

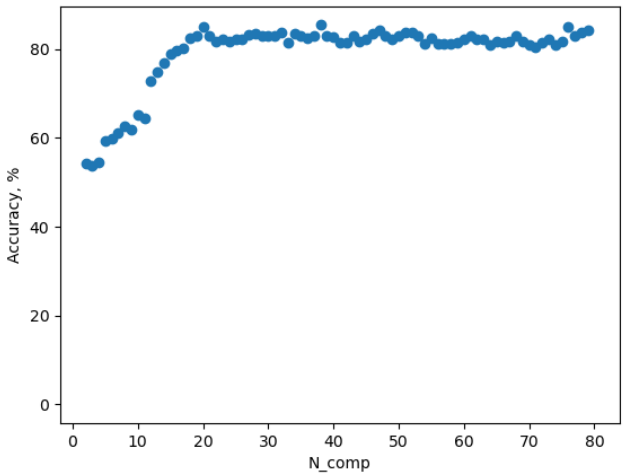


Figure 3.9. Dependence of the classification accuracy on the number of principal components used.

It can be seen that when the number of principal components used for classification is higher than 20, the accuracy of the multiclass classification reaches a ceiling value of about 90%. This correlation was the same for all groups of spectra, but the number of principal components used for classification, for which the max constant value of accuracy was reached, varied between 20 and 30. In addition, no improvement in efficiency of classification was achieved by varying the kernels. Therefore, the principal component analysis was implemented using linear kernel in all cases.

The use of different kernels for the SVM also provided very similar results, so the kernels that required less computation time were used for spectrum classification (linear kernel and RBF kernel). A similar approach was used to choose the optimal hyperparameters for the classifiers used.

The last step of optimization was to find the optimal cross-validation method (LOOCV/*k*-fold CV) and the optimal value of the *k* parameter (for *k*-fold CV) which provided an optimal balance between the resulting accuracy and computation time. For this purpose, the value of parameter k was varied from 4 to 20 for all groups of spectra. SVM (with linear and RBF kernels) and LDA were used as classifiers, with only the most accurate result selected for further analysis. The number of principal components used after PCA was equal to 25 in all cases. For almost all groups of spectra, the difference between the minimum and maximum values of the multiclass classification accuracy was less than 2.5%. For further calculations, the value of *k* was chosen to be 8, since the classification results corresponding to this value were maximal or close to it for all groups of spectra. In addition, it does not require a long computation time, compared to LOOCV.

The results of the selection are presented in **Table 3.5**. The use of more than one hyperparameter value/option in the table means that the values of multiclass classification accuracy obtained in the corresponding cases differed by no more than 1%, and the required computation time differed by no more than 10 seconds.

Table 3.5. Selection of hyperparameters for methods used in preclinical study.

Name of the method	List of optimized hyperparameters	Selected hyperparameters
Principal component analysis (PCA)	Number of PCs	25
	Kernel	Linear
Support vector machine (SVM)	Kernel	Linear/ RBF
	Regularization parameter value	1
	Decision function type	OVO
Multi-layered perceptron (MLP)	Number of hidden layers	1
	Size of each hidden layer	80
	Type of the activation function	ReLU
Linear discriminant analysis (LDA)	Type of solver	SVD
Random forest (RF)	Number of trees	100

3.3.2. Results of classification for different groups of spectra

The classification of each of the 24 groups of spectra was done by all methods in parallel and a classifier with the best performance were chosen for each group of spectra. All calculations were repeated 5 times to get statistically relevant results. Average accuracy values obtained by the different classifiers (SVM, MLP, DA and RBF) for all modalities (DR and all AF) and all SDS are shown on **Figure 3.10** (for sake of clarity, error bars are plotted only for standard deviation values greater than 1%).

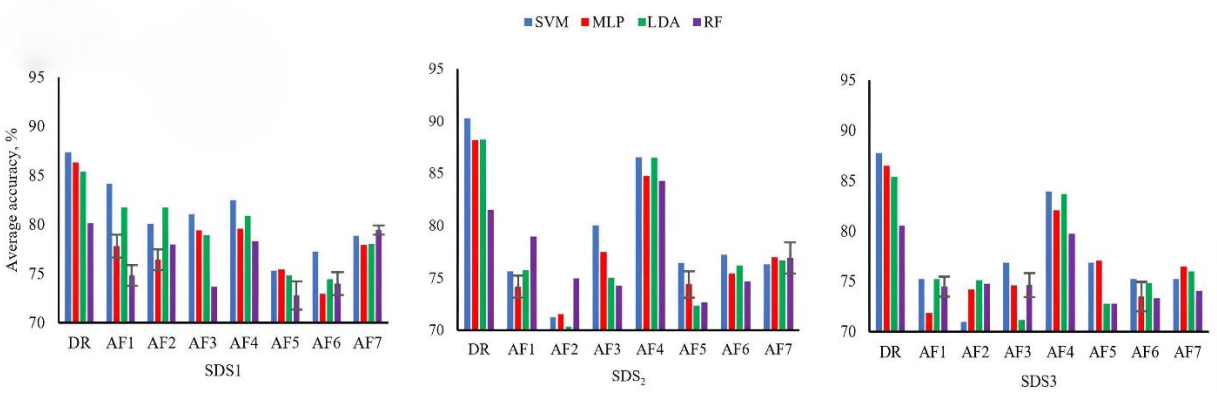


Figure 3.10. Average classification accuracy obtained by different classifiers (SVM, MLP, LDA and RBF) for all modalities (DR and all AF) and all SDS; error bars indicate standard deviation (SD) values corresponding to five repeated calculations for each combination of source, detector and classifier (error bars not shown for SD < 1%).

It can be seen that all the classification methods used, except RF, showed very similar accuracy in most cases. It can also be noted that for all SDS, the best results were obtained for the classification of DR spectra. The multiclass classification accuracy and percentage of potentially precancerous lesions classified as healthy are presented in **Table 3.6** for all groups of spectra.

Table 3.6. Values of accuracy and misclassification for PP states for all modalities and sub-modalities.

Group of spectra	**SDS_1 = 271 µm**		**SDS_2 = 536 µm**		**SDS_3 = 834 µm**	
	Acc., %	**Err., %**	**Acc., %**	**Err., %**	**Acc., %**	**Err., %**
DR	87.40	4.35	**90.24**	**0.87**	86.99	6.09
AF_1	**84.15**	**6.96**	78.05	12.17	76.42	14.78
AF_2	81.71	6.96	75.20	14.78	77.24	21.24
AF_3	81.30	6.96	78.05	10.43	77.64	13.04
AF_4	82.52	11.30	**86.59**	**6.09**	83.33	7.73
AF_5	77.24	12.17	78.46	13.91	76.02	11.30
AF_6	75.61	19.13	77.64	11.30	73.98	24.35
AF_7	79.27	13.91	79.27	13.91	77.24	19.13

For a more detailed analysis, two best results were selected for each SDS. The corresponding confusion matrices are presented in **Table 3.7**.

Table 3.7. Confusion matrices corresponding to the most accurate results for every SDS.

SDS_1 = 271 μm									
	DR					**AF_1**			
	AH	**CH**	**D**	**H**		**AH**	**CH**	**D**	**H**
AH	71.19	0	23.73	5.08	**AH**	66.10	0	22.03	11.86
CH	2.13	95.74	2.13	0	**CH**	0	100	0	0
D	17.86	0	78.57	3.57	**D**	28.57	0	69.64	1.79
H	0	0	0	100	**H**	2.38	0	0	97.62
SDS_2 = 536 μm									
	DR					**AF_4**			
	AH	**CH**	**D**	**H**		**AH**	**CH**	**D**	**H**
AH	88.14	0	10.17	1.69	**AH**	74.58	0	16.95	8.47
CH	0	93.62	2.13	4.26	**CH**	0	100	0	0
D	21.43	0	78.57	0	**D**	19.64	0	76.79	3.57
H	2.38	0	0	97.62	**H**	2.38	0	3.57	94.05
SDS_3 = 834 μm									
	DR					**AF_4**			
DR	**AH**	**CH**	**D**	**H**		**AH**	**CH**	**D**	**H**
AH	67.80	0	23.73	8.47	AH	66.10	0	20.34	13.56
CH	0	100	0	0	CH	2.13	95.74	2.13	0
D	17.86	0	78.57	3.57	D	26.79	0	71.43	1.79
H	0	0	1.19	98.81	H	3.57	0	0	96.43

It can be seen that for all three values of SDS the highest multiclass classification accuracy (90,24%) and lowest value of error in classification of PP states (0,87%) were obtained for DR spectra, which is a significant improvement compared to the results obtained in the previous studies. This may be explained by the fact that the implementation of PCA for feature extraction without preliminary step of feature selection provides greater efficiency in extracting important information from the initial DR spectra which is also indirectly confirmed by the fact that the classification accuracy of the autofluorescence spectra has also increased compared to the same studies. According to histological criteria, higher classification accuracy of DR spectra may relate both to a change in the absorption properties of the skin with PP states due to keratosis, and scattering properties of the skin with PP lesions, due to disorganisation of the cell layers.

Among the different SDS values, the best results for AF spectra (Acc = 84,15%, Err = 6,96%) with excitation wavelengths $\lambda < 400$ nm were obtained for SDS_1 = 271 μm. This may be due to the fact that these wavelengths have a lower penetration depth into the skin, which affects the quality of the spectra obtained for larger SDS values. In addition, according to histological classification criteria, most of the changes in skin properties due to lesions are related with changes in the concentration of keratin, which has a fluorescence excitation peak in the 380-400 nm range [170]. Also in cases of dysplasia, degradation of collagen and elastin

in the dermis can affect the final accuracy of the classification, as collagen, elastin and their cross-links also have fluorescence excitation peaks in the 360-400 nm range.

However, the best results for DR spectra (Acc = 90,24%, Err = 0,87%) and AF spectra with an excitation wavelength of $\lambda \geq 400$ nm (Acc = 86,59%, Err = 6,09%) were obtained for SDS_2 = 536 μm. It can be explained by the fact that this value of SDS provides the optimal balance between the penetration depth into the skin and the proportion of noise in the signal, as increasing SDS increases the contribution of the signal received from the deeper layers of the skin (including the dermis) to the resulting spectrum. Moreover, the results obtained for the DR spectra collected at SDS_2 are the best of all the results obtained in this step. Consequently, this suggests significantly changes of scattering properties of skin layers (including the dermis) related to their morphological and structural modifications.

It can also be noticed that in all cases, the precision values of classification of benign classes (H and CH) ranges from 93 to 100%, while the precision values of classification of PP states ranges from 66 to 88%. At the same time for the best result presented (DR spectra for SDS_2), only 1.69% of all samples corresponding to atypical hyperplasia were classified as healthy, whereas for dysplasia there were no such samples at all. Thus, the number of samples corresponding to potentially precancerous skin states that were classified as benign or healthy was 0.8% of all samples that belong to PP classes according to the reference classification. In order to improve this result, all spectra were re-divided into two groups, PP and benign and then spectra processing pipeline was repeated. The results presented in **Table 3.8**. All results presented are consistent with the classification of the DR spectra, as in the case of multiclass classification, the best results were obtained for it among both modalities and all groups of spectra.

Table 3.8. Results of mice skin spectra binary classification obtained using DR spectra for every SDS.

SDS_1		
	Benign	**PP**
Benign	93.13	6.87
PP	0	100

SDS_2		
	Benign	**PP**
Benign	95.42	4.58
PP	1.74	98.26

SDS_3		
	Benign	**PP**
Benign	96.18	3.82
PP	1.74	98.26

It can be seen that in case of binary classification for spectra obtained at SDS_1 the proportion of PP classified as benign is 0. For SDS_2 and SDS_3 the obtained coefficients of the confusion matrices are very similar.

3.3.3. Results of Data fusion

The next step was to combine the results from the different groups of spectra by fusion of decisions in several ways: fusion of pre-processed spectra by linear concatenation, fusion of features after PCA, fusion of decisions for all groups of spectra, fusion of decisions according to the initial spectra modality or excitation wavelength value (for AF spectra), and fusion of decision according to SDS of the initial spectra.

Fusion of initial data and fusion of features include only one stage of classification, so it makes sense to compare the results obtained using such strategies with the results of single-modality classification shown in **Table 3.6**. (**Section 3.2.2**). The results obtained after applying these two strategies for all listed options of fusion of spectra by groups are presented in **Table 3.9**.

Table 3.9. Best results obtained by using fusion of initial data and fusion of features for modalities/sub-modalities.

Groups of spectra for fusion	**Fusion of initial data**		**Fusion of features**	
	Acc, %	**Err, %**	**Acc, %**	**Err, %**
All groups	85,81	6,26	86,94	3,81
DR	85,42	7,36	86,17	4,22
AF_1	82,51	4,85	80,47	7,26
AF_2	74,01	2,59	80,94	7,02
AF_3	76,84	8,00	77,23	4,87
AF_4	76,49	6,67	79,30	5,72
AF_5	78,09	5,63	71,91	5,08
AF_6	75,22	5,04	75,59	8,16
AF_7	72,35	4,02	72,02	8,97
SDS_1	85,79	3,57	82,07	5,35
SDS_2	85,81	4,17	82,54	6,98
SDS_3	82,55	7,84	84,15	4,41

When comparing the results obtained by fusion of initial data by concatenation and fusion of PCA features with the results presented in **Table 3.6**, it can be concluded that the use of either of these data fusion strategies does not lead to improved classification results. As a consequence, a detailed analysis of these results, as well as the application of these methods to the analysis of spectra measured on human skin, is meaningless.

The next step was to test the effectiveness of the decision fusion methods such as Majority Voting, Stacking and Weighted Voting for each group of spectra. All corresponding results are presented in **Tables 3.10** and **3.11** respectively.

Table 3.10. Results obtained by using decision fusion for modalities/sub-modalities.

Groups of spectra for fusion	Majority voting		Stacking		Weighted Voting	
	Acc, %	Err, %	Acc, %	Err, %	Acc, %	Err, %
All groups	**89.02**	**3.48**	86.59	7.83	-	-
DR	89.84	1.74	**90.24**	1.74	**90.24**	**0.87**
AF_1	80.49	**4.35**	84.15	8.70	**84.96**	6.96
AF_2	80.08	8.70	80.08	15.65	**81.71**	**6.96**
AF_3	79.27	**4.35**	82.52	12.71	**83.33**	9.57
AF_4	86.99	6.96	87.80	7.83	**89.02**	**5.22**
AF_5	76.02	10.43	80.08	12.71	**83.30**	**9.57**
AF_6	73.58	17.39	79.67	17.39	**80.08**	**14.78**
AF_7	78.46	15.65	80.49	19.13	**82.11**	**13.04**
DR and all AF (step 2)	87.80	4.35	85.77	11.30	**93.90**	**1.74**

Table 3.11: Results of decision fusion by SDS.

Distance	Majority voting		Stacking		Weighted Voting	
	Acc, %	Err, %	Acc, %	Err, %	Acc, %	Err, %
SDS_1	86.18	7.83	86.18	**3.48**	**90.24**	5.22
SDS_2	88.21	10.43	92.68	1.74	**94.31**	**0.87**
SDS_3	82.52	13.91	85.37	6.96	**89.43**	**6.09**
Final 2	86.59	9.57	91.06	4.35	**94.31**	**0.87**

It can be noted that, in most cases, implementation of WV provides the best results among all three methods for both evaluated parameters. It can be explained by the influence of overfitting in case of stacking and the lack of corrections for the difference in classification efficiency based on different groups of spectra in case of MV. In addition, as in the previous case, the selection of the best result was based on the multiclass classification accuracy. However, the use of WV requires a much longer computation time.

Figure 3.11 shows the average accuracy ± SD bar graphs allowing to compare the performance of the different classifiers (SVM, MLP, LDA and RF) for different strategy of combining/choosing data for decision fusion by stacking. It can be seen that Random Forest and SVM provided the best performance in most cases.

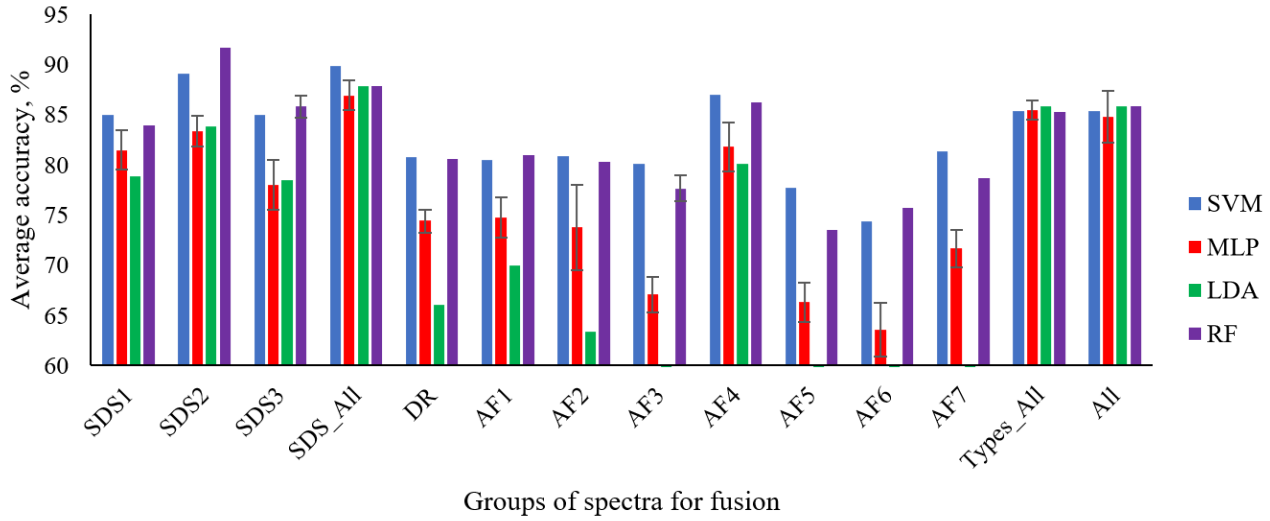


Figure 3.11. Average accuracy for stacking (decision fusion) obtained by different classifiers (SVM, MLP, DA and RBF) for all ways to combine data (by SDS, by modality and all data at once); error bars indicate SD values corresponding to five repeated calculations for each combination of source, detector and classifier (error bars not shown for SD < 1%).

The best results of decision fusion, as for the separate classification, were obtained for DR spectra and AF spectra with excitation wavelength of $\lambda = 400$ nm. In case of the DR spectra, the confusion matrix coefficients were identical to the ones presented in Table 4(c), which correspond to the classification results based on DR spectra for $SDS_2 = 536\mu m$. The use of decision fusion for the AF spectra gives an average improvement of 3-5% in multiclass classification accuracy.

The best results for fusion of decisions by group of spectra, that were obtained by using two steps of weighted voting are presented in **Table 3.11**. It can be seen that the precision values for all classes increased by 5-7% and reached 100% for benign states compared to the results presented in **Table 3.7**. At the same time, the error in classification of atypical hyperplasia as healthy tissue also increased. This was due to the fact that the selection of the best combination of coefficients for WV was based on the multiclass classification accuracy.

When combining the SDS classification results, the best results (Acc = 94,31%, Err = 0,87%) were obtained for $SDS_2 = 536\mu m$, as in the previous case. The corresponding results that were obtained by using WV are presented in **Table 3.12**.

Table 3.12: Classification results (confusion matrix) based on decisions obtained from all types of spectra.

	AH	**CH**	**D**	**H**
AH	91.53	0	6.78	1.69
CH	0	100	0	0
D	16.07	0	83.93	0
H	0	0	0	100

Compared to the results presented in **Table 3.7**, the value of precision for all classes also increased by 5-7% and reached 100% for benign states, as in the case of decision fusion by modality. The use of WV and stacking for decision fusion by SDS value increased the multiclass classification accuracy by 2-4%. At the same time, additional step of decision fusion produced no improvement with respect to the result obtained for $SDS_2 = 536\mu m$.

For the case of binary classification, the fusion of decisions was done in the same way as for multi-class classification. The best results of fusion (obtained at the second decision fusion step) are presented in **Table 3.13** for distances (a) and groups of spectra (b).

Table 3.13: Results of decision fusion for binary classification.

SDS_1		
	Benign	**PP**
Benign	100	0
PP	0.87	99.13

(a)

SDS_3		
	Benign	**PP**
Benign	99.24	0.76
PP	0.87	99.13

(b)

It can be noted that the percentage of misclassifying the PP state as healthy is less than 1%. Moreover, the resulting error is most likely the result of statistical error due to the use of cross-validation. At the same time, the corresponding value of precision, compared to the results obtained without decision fusion, has increased and is close to 100%.

3.4. Conclusion

Chapter three discussed the results of applying data processing methods, including machine learning and data fusion techniques, to analyse bimodal spectroscopic data obtained *in vivo* from the skin of mice with benign and precancerous skin lesions.

The spectroscopic device used to measure the spectra included a system of linearly variable band-pass filters that, in combination with a xenon lamp, were used to obtain DR and multi-excitation AF spectra. The optical probe of the device contained several detecting fibres allowing to measure spectra for different distances between the source and the detector, which allowed to vary the proportion of detected photons with a greater depth of penetration into the

skin. According to the combination of source and SDS for which spectra were obtained, all of them were categorised into 24 groups.

The spectra corresponding to each group were processed separately, followed by a one- or two-step data fusion. The feature extraction was made by using principal component analysis, after which the spectra were classified using Support Vector Machine, Multi-Layer Perceptron, Linear Discriminant Analysis and Random Forest. The next step was to use data fusion, which was done in three ways: combining the initial data before feature extraction step, combining PCA features obtained from each group of spectra and combining the results of classification obtained for each group of spectra by majority voting, stacking or voting with optimized weights for each group.

The highest accuracy of the multi-class classification was achieved using support vector machine for the first step of classification and weighted voting to combine all types of spectra corresponding to $SDS_2 = 536\mu m$ for the second step of classification and was 94.41%. It can be explained by the fact that this value of SDS provides an optimal balance between depth of penetration into the skin and the proportion of noise in the signal. In addition, the highest accuracy obtained in binary classification of all samples into benign and potentially precancerous conditions after applying all data processing, classification and decision fusion methods was almost 100%.

Among the different spectral types, the highest classification accuracy was obtained for the diffuse reflectance spectra for all SDS. This observation can be explained by the fact that for the whole spectral range the diffuse reflectance spectra are equally affected by the absorption properties of internal skin components range as well as by scattering properties, which are influenced by pathological states of the skin. However, this fact cannot be clearly confirmed by the list of methods used in this chapter.

Thus, the combination of experimental techniques and methods of data processing and analysis allowed to obtain accuracy values for diagnosing benign and precancerous skin conditions of mice over 90%, which indicates its high efficiency, as well as the prospect of its application for obtaining and analysing spectra obtained from human skin.

Chapter 4. Clinical study on human skin

4.1. Introduction

Once the methodology for analysing spectra obtained in the preclinical study on mouse skin had been selected, it was planned to apply it to spectra obtained in the clinical study on human skin. However, the pilot device used in the preclinical study had major differences from the SpectroLive Device used in the clinical study, so all spectra preprocessing methods, including smoothing and noise removal methods, had to be tested again. Also, due to the fact that many classification and data fusion methods showed similar performance in the preclinical study, it was decided to test all possible combinations of such methods.

In addition to the differences in experimental method between the preclinical and clinical study, it should also be mentioned that the level of diversity in the properties and condition of human skin samples is much higher than in laboratory mice, which may negatively affect the results of spectra classification obtained in the clinical study. A comparative analysis of the results of the preclinical and clinical studies will be given in **Section 4.3.4**.

This chapter will describe the materials and methods of the clinical study as well as the results obtained. Particular attention will be paid to comparing the performance of different modalities as well as comparison of the results obtained both with the results of the preclinical study described in Chapter 3 and with the results obtained by NMSC diagnostic studies using unimodal and multimodal optical methods described in Chapter 1.

4.2. Materials and methods

4.2.1. General information about clinical study

The clinical study was conducted by a plastic surgeon at the Metz-Thionville Regional Hospital (France) and approved by both the French National Drug Agency (ANSM) and an ethical committee (CPP Est III). Between years 2017 and 2021, the clinical trial named "SpectroLive" included 131 patients with one or more lesions supposed to be a skin carcinoma. A brief description of the clinical study is provided hereafter but detailed materials and methods information can be found in [44].

It is also important to note that the set of obtained spectra will be described in 4.2.2.1 and all additional information are publicly available in a database format and can be used to train statistical models for the diagnosis of skin carcinomas as well as to determine its parameters contained in the database.

4.2.2. Instrumentation and measurement procedure

The instrumentation used in this clinical study was the patented spatially-resolved DR and multiply excited AF spectroscopy device SpectroLive [41]. A simplified scheme of the SpectroLive device is provided in **Figure 4.1** (detailed technical and metrological information can be found in [40]).

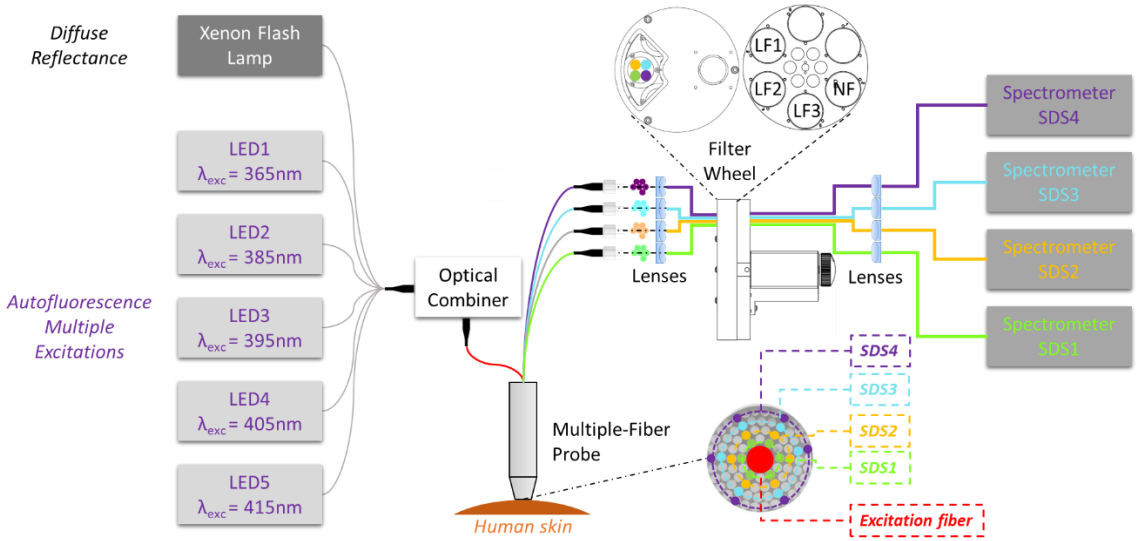


Figure. 4.1 Simplified schematic representation of the SpectroLive medical device including a Xenon lamp wide-band source for DR measurements and five LED narrow-band sources for AF excitations, the multiple-fiber optics probe featuring 4 Source-Detector Separations (SDS1-4), a filter wheel equipped with long-pass emission filters for AF measurements and 4 spectrometers.

The device is equipped with five high power fiber-coupled LED modules (Mightex®) LED1-5 in combination with in-line band pass filters to provide narrow-band excitation peaks between 365 and 415 nm (see details in Table 4.1). A Xenon flash lamp from Hamamatsu (Massy, France) was used as wide-band light source to obtain diffuse reflectance spectra (see again Table 1). It has to be noticed that a Lambertian surface spectralon target (SRS-99-010, Labsphere®, Poynton, Great Britain) with spectrally flat response in the wavelength range of acquisition was used to obtain the reference signal for DR spectra measurements.

The optical probe has a measuring area diameter of 3.5 millimetres and contains one central 600 µm-diameter fiber for illuminating the skin tissue and 4 groups of 6 200µm-diameter fibers each, corresponding to 4 Source-Detector Separations (SDS1-4) from 400 to 1000 µm between the centres of the light emitting and collecting fibers. The light collected by

each group of 6 fibers is transmitted to 4 spectrometers (Mightex®, Pleasanton, CA, USA), respectively. On the way of the light a filter wheel allows, before the spectrometers, for selecting appropriate long pass AF emission filters providing high-blocking (optical density greater than 6) of the AF excitation peaks. The main characteristics of the device are presented in **Table 4.1**.

Table 4.1. Main light source and detection characteristics of the SpectroLive device (*FWHM = Full Width at Half Maximum).

Modality	**Source wavelength (nm)**	**Integration time (s)**	**Acquisition wavelength range (nm)**	**Source-Detector Separation (SDS) (µm)**
Diffuse reflectance spectroscopy	Range: 350-800	0,05	340-785	SDS_1 = 400 SDS_2 = 600 SDS_3 = 800 SDS_4 = 1000
Multi-excitation Autofluorescence spectroscopy	Peak (FWHM*): AF_1 = 365 (4) AF_2 = 385 (9) AF_3 = 395 (7) AF_4 = 400 (9) AF_5 = 415 (10)	1		

A “single sequence” measurement lasts six seconds and consists of the acquisition of 24 spectra including 20 AF spectra (captured for 5 excitation peaks at 4 SDS), and 4 DR spectra (1 source signal and measurement at 4 SDS). In practice, this single measurement was repeated three times for each tissue point (without moving the probe) in order to provide “single point” measurements with an increased signal-to-noise ratio or reduced influence of random outliers in the resulting spectra (see **Section 2.2.2**) the. The integration time, chosen to have an optimal compromise between signal-to-noise ratio and clinical ergonomics, was one second for each of the autofluorescence spectra and 50ms for the diffuse reflectance spectra [40]. In addition, dark noise was also measured during each measurement to account for the influence of thermally generated electrons.

The SpectroLive device is essentially bimodal in that it provides both AF and DR modalities, however, it features 5 sources of AF excitation. As a consequence, each of the sources of excitation of autofluorescence will be considered as an independent modality for data fusion steps in our classification pipeline to simplify the description of the methods used and the results obtained.

4.2.3. Clinical and spectroscopic dataset

The total number of lesions for 131 patients participated in the study was 218. The age of all patients, their gender, their skin phototype and the anatomical sites where the lesions were located are presented in **Table 4.2**. Extended information about the dataset can be found in [44].

Table 4.2. Distribution (%) of the 131 patients in classes of gender, age, phototype and lesion anatomical site (sometimes several lesions per patients).

Patient parameters		**Distribution (%)**
Gender (131 patient)	Male	69
	Female	31
Age (131 patient)	30-50	2
	50-60	6
	60-70	18
	70-80	27
	80-90	37
	90-100	11
Phototype (131 patient)	I	13
	II	65
	III	20
	IV	2
Anatomical site (218 lesions for 131 patients)	Forehead, vertex, ear, temple, eyelid	55
	Cheek, Nasolabial fold	23
	Neck, nape, shoulders	10
	Back	4
	Arm, hand	4
	Legs	4

For each suspected lesion, five different locations were defined for single point spectroscopic measurements (see **Figure 4.2 (a)**): one at the "Lesional" area itself (L), two at a short "Peri-Lesion" distance from the lesion (PL), and two more at the edge of this safety margin area marked under the guidance of a plastic surgeon which is supposed as "Non Lesion" (NL) healthy skin. After spectroscopic acquisitions, suspicious lesions were excised along the boundaries of the marked area for the sake of a clean surgical suture. In addition to the NL spectra measured near the lesion, spectra were also measured on several other reference healthy skin in the wrist region for all patients, as shown in **Figure 4.2 (b)**, in order to be further exploited for standardization steps in the data processing pipeline.

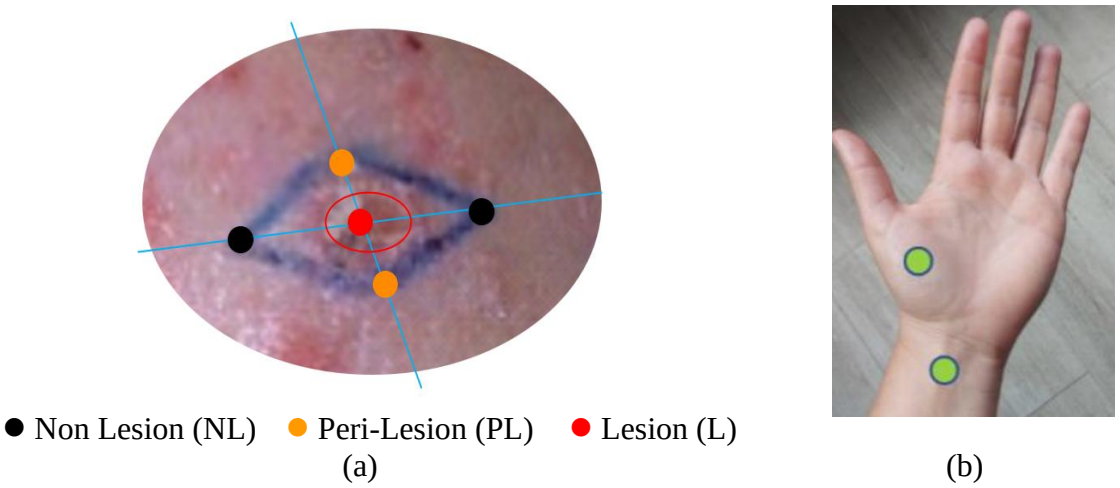


Figure 4.2. Schematic representation of the spectroscopic measurement spots on: (left) Lesion (L), Peri-Lesion (PL) and Non Lesion (NL) areas in and around the lesion area (red circle) and along the short and long axes (blue lines) of the blue ink delimited surgical spindle; (right) healthy areas on palmar hand and inner wrist.

Based on the results of histopathological analyses acting as ground truth, each L, PL and NL areas of the 218 anatomical sites (851 measured areas in total) were assigned to one of the four main following diagnostic classes: Basal Cell Carcinoma (BCC, number of measured areas n = 90), Squamous Cell Carcinoma (SCC, n = 57), Actinic Keratosis (AK, n = 96) or Healthy skin (H, n = 608). The average spectra for all four main diagnostic classes, with their corresponding standard deviations, are shown in **Figures 4.3-4.6**.

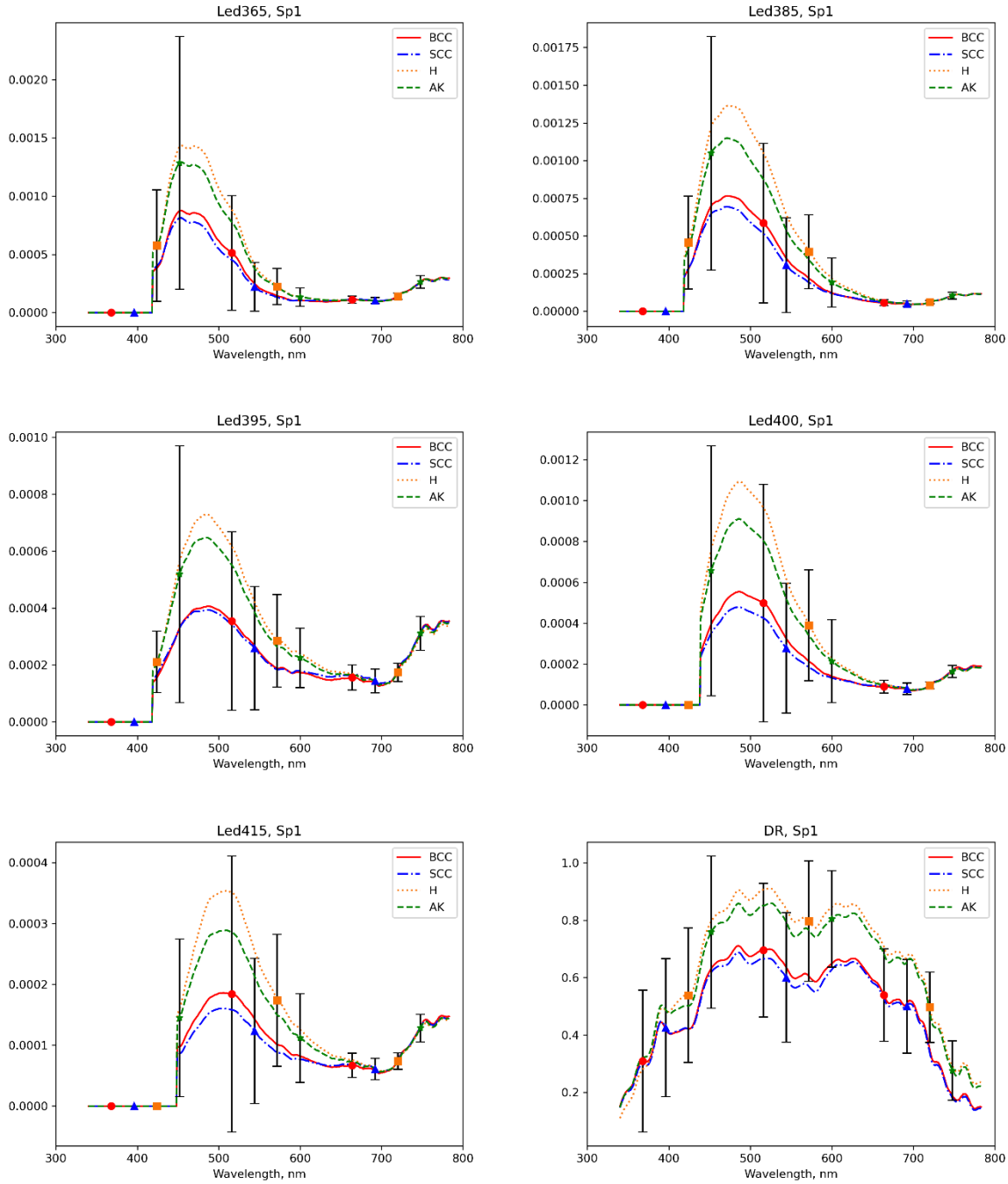


Figure 4.3. Mean spectra for main diagnostic classes obtained at SDS_1 = 400μm.

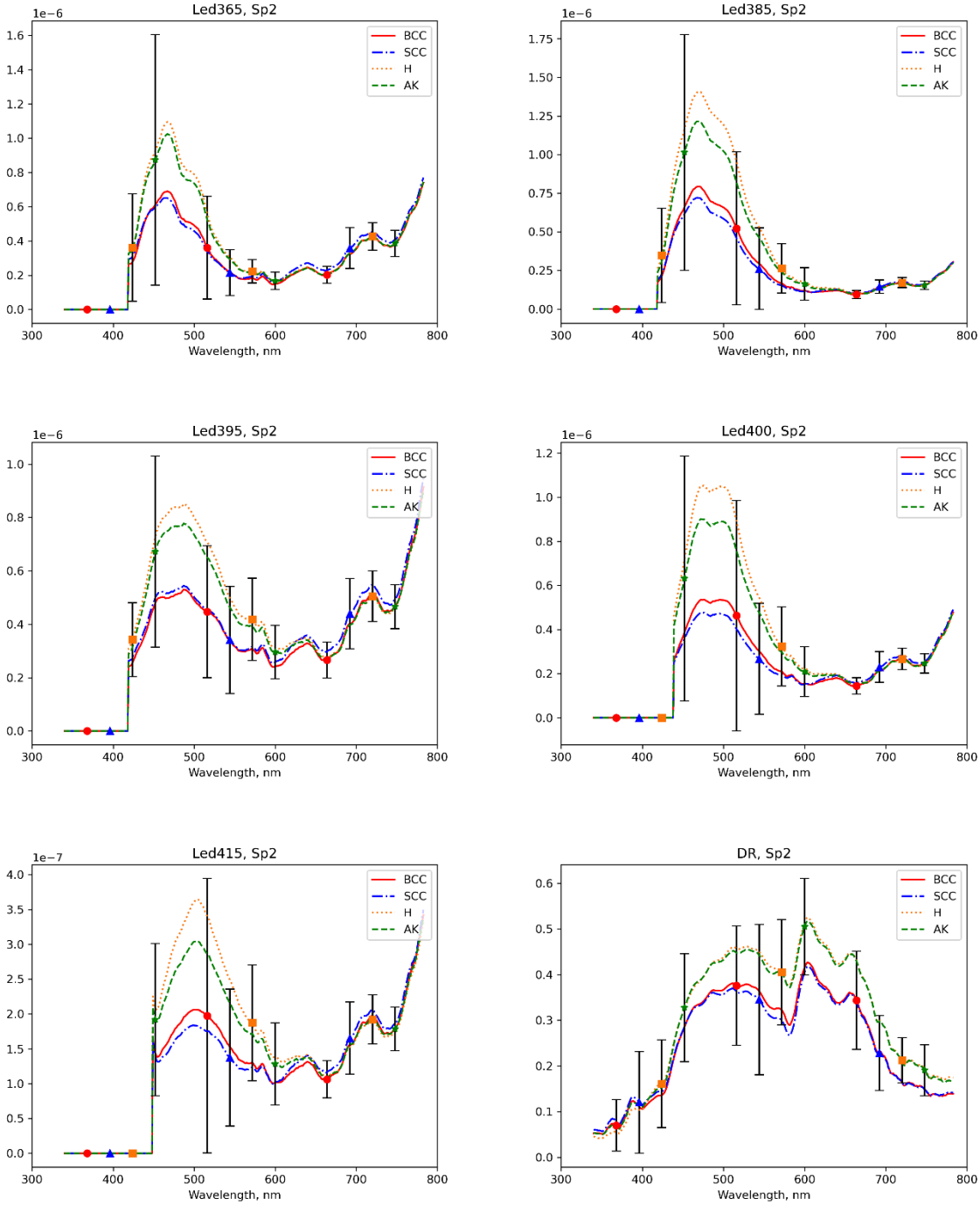


Figure 4.4. Mean spectra for main diagnostic classes obtained at SDS_2 = 600μm.

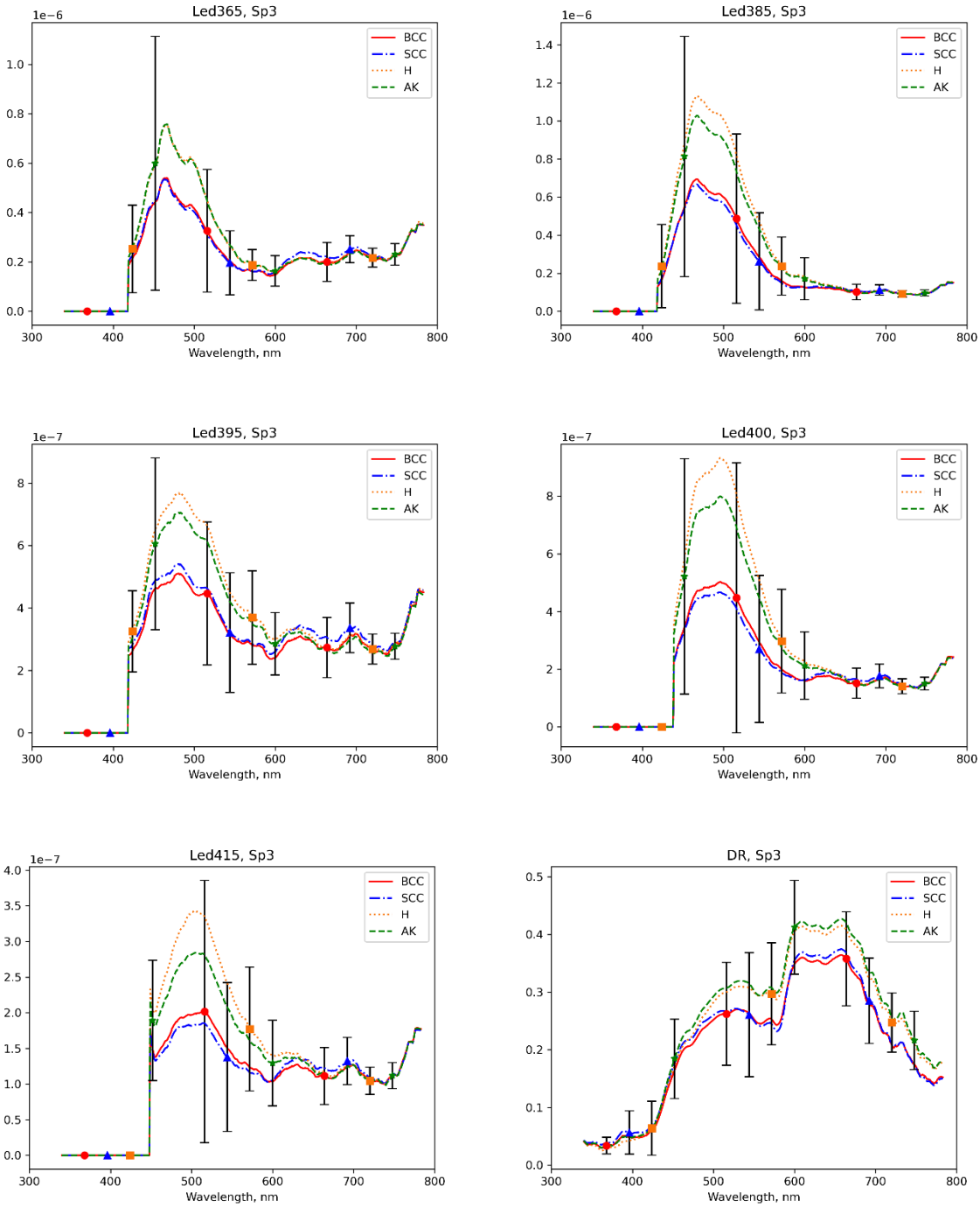


Figure 4.5. Mean spectra for main diagnostic classes obtained at SDS_3 = 800μm.

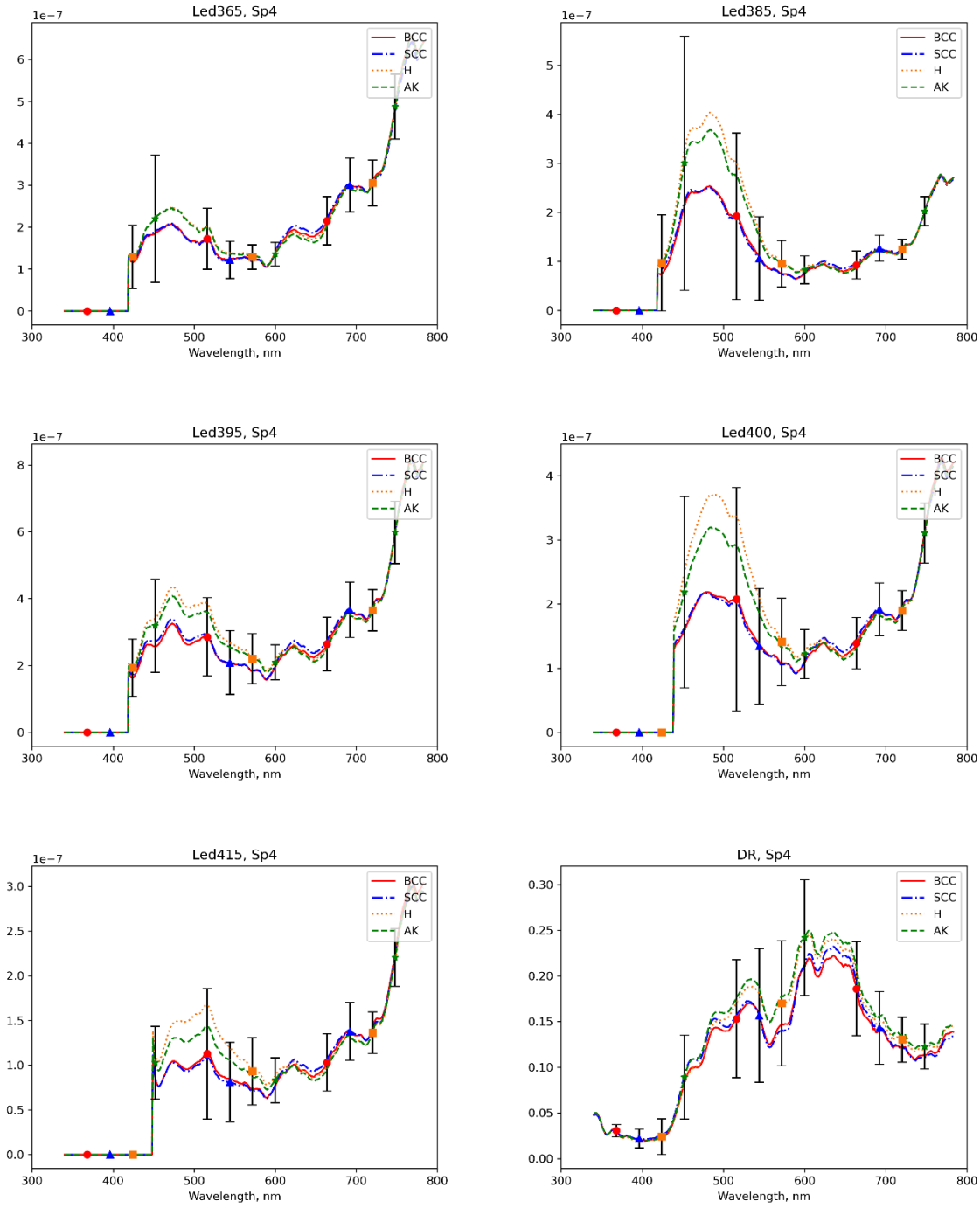


Figure 4.6. Mean spectra for main diagnostic classes obtained at SDS_4 = 1000μm.

It can be noticed in **Figures 4.3-4.6** that the Standard Deviation bars for each class, representing the intraclass variability in the spectral intensity curve shapes, is relatively large (as well as in a preclinical study on mouse skin), so that the amplitude ranges of spectra corresponding to each class overlap significantly and degrades the interclass separability. In addition, most average AF spectra contain several separated fluorescence peaks, which is related to the fact that the shapes of the spectral curves can vary significantly within the same

class due to the high level of variability in the skin properties of the patients involved in the study. Thus, a careful optimization of every step of the data processing and classification pipeline is extremely important.

4.2.4. Spectroscopic data processing pipeline

An overview of the general scheme of the proposed data processing pipeline is given in **Figure 4.7**. Th pipeline includes: (I) raw spectra filtering and correction, (II) detection of measurements that contains outliers, (III) normalization, (IV) dimensionality reduction and (V) classification.

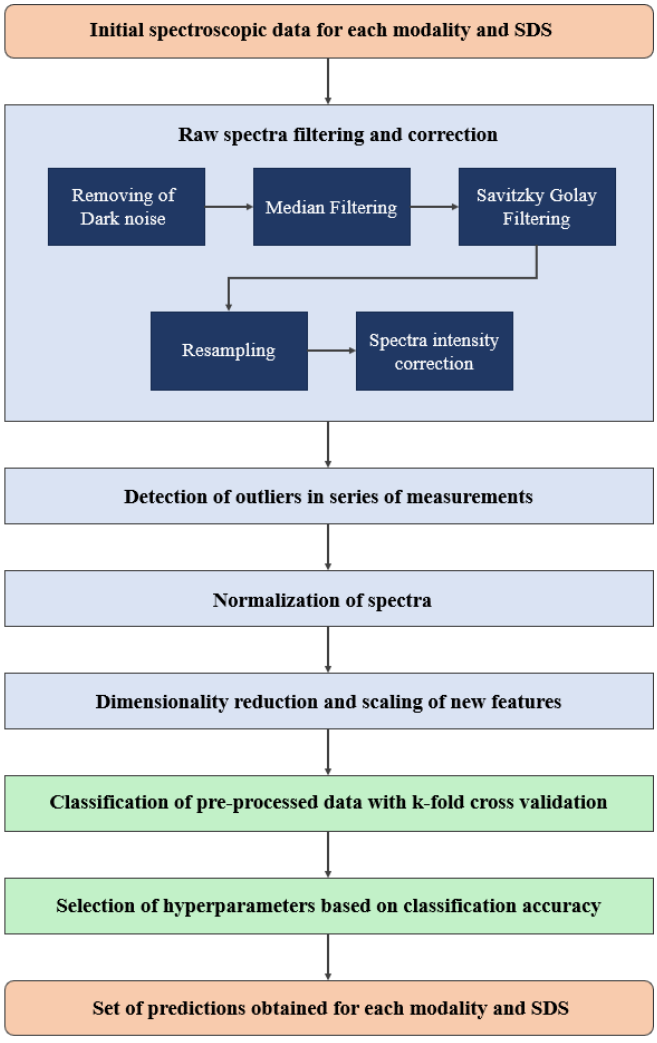


Figure 4.7. Spectroscopic data processing pipeline diagram overview.

4.2.4.1. Raw spectra filtering and correction

Pre-processing of the raw spectra of all modalities included median filtering to suppress impulse noise (window size = 3), dark noise subtraction to account for the influence of thermally generated electrons in spectrometers, smoothing using Savitzky-Golay filter (window

size =81, polynomial order = 4) and resampling (from 0.3 nm step to 1 nm step). For the autofluorescence spectra, intensity correction of the spectral response was also performed for each detection channel to remove the influence of optical system components on the obtained spectra.

4.2.4.2. Detection of outliers in series of measurements

As mentioned in **Section 4.2.1.2**, one tissue point measurement consists in 3 successive measurements at the same tissue point (without moving the probe). This was exploited to detect and remove any potential "outlier" signal for each series of three measurements before averaging. This outlier rejection is the same for all the modalities and SDS, and is based on relative dispersion in the form of average variation coefficients calculated for DR and AF spectra respectively as:

$$\overline{VC}_{AF} = \frac{1}{N} \cdot \sum_{\lambda_i=\lambda_1}^{\lambda_2} \frac{\sqrt{\frac{\sum_{j=1}^{3}\left[AF_j(\lambda_i) - \overline{AF}(\lambda_i)\right]^2}{3}}}{\overline{AF}(\lambda_i)} \quad (4.1)$$

$$\overline{VC}_{DR} = \frac{1}{N} \cdot \sum_{\lambda_i=\lambda_1=340\ nm}^{\lambda_2=780\ nm} \frac{\sqrt{\frac{\sum_{j=1}^{3}\left[DR_j(\lambda_i) - \overline{DR}(\lambda_i)\right]^2}{3}}}{\overline{DR}(\lambda_i)} \quad (4.2)$$

where $AF(\lambda)$ and $DR(\lambda)$ stand for the AF and DR spectra, while $\overline{AF}(\lambda)$ and $\overline{DR}(\lambda)$ correspond to the average spectra for the three measurements of each point on skin, $\lambda_1\ and\ \lambda_2$ define the boundaries of the wavelength range used for the calculation, *N* is the number of points in the wavelength range of $[\lambda_1; \lambda_2]$.

Variation coefficients $\overline{VC}_{AF}$ and $\overline{VC}_{DR}$ are essentially averaged standard deviations normalized to the mean at each wavelength. Consequently such coefficients can be used as an independent characteristic of the "stability" (outlier presence or non-constant spectrum intensities) of a variable. On the contrary, the use of the standard deviation without considering the mean value of the variable makes no sense since the maximum and minimum intensity values in the same spectrum can differ by orders of magnitude. However, since spectra contain hundreds of points it is more appropriate to use values averaged over the spectral range instead of exploiting the coefficient of variation for a single point.

The limits of the wavelength range for calculating $\overline{VC}_{AF}$, are given in **Table 4.3** according to the excitation wavelength. Lower bounds (λ_1) correspond to the shortest wavelength in the transmitted band of the fluorescence emission filters, while higher bounds

(λ_2) were defined as the longest wavelength beyond which significant increase of the noise level in the spectra was observed.

Table 4.3. Wavelength ranges [λ_1:λ_2] for the calculation of the average variation coefficients of AF spectra.

Modality	λ_1, nm	λ_2, nm
AF_1, AF_2 and AF_3	421	550
AF_4 and AF_5	451	580

Checking for the presence of outliers for each series of three measurements for the same skin area before averaging was performed separately for each modality and SDS and consisted of two stages. At the first stage, the mean variation coefficients for the series of three measurements was calculated and, if its value did not exceed 0.08, the series of measurements was considered to contain no outliers and all three spectra were used for subsequent averaging. In the opposite case, a separate variation coefficient was calculated for each of the three possible pairs of measured spectra. The main idea of this approach was to assume that if two of the three spectra in a series of measurements were slightly different from each other and significantly different from the third, the variation coefficient for the corresponding pair of spectra should be smaller than the variation coefficients for the other two pairs of spectra. Thus, if one of the coefficients of variation calculated for pairs of spectra was less than 0.08, the corresponding pair of spectra was considered correctly measured and used for subsequent averaging. At the same time, if all coefficients of variation calculated for the three pairs of spectra exceeded 0.08, the measurement was considered as incorrect and was removed from subsequent analyses for the corresponding combination of modality and SDS.

Indeed, the threshold value calculated on trios and pairs of spectra was determined as a trade-off between (i) identifying and excluding measurements in which one or all three spectra were significantly different from the others and (ii) limiting the number of spectra removed from the dataset. The limit value of 0.08 was, on the one hand, small enough to filter out outliers in the measurements and, on the other hand, allowed to keep on average more than 85% of the spectra corresponding to non-healthy classes in the analyses for each modality and SDS. The general scheme of filtering of outliers in measurements is shown in **Figure 4.8**.

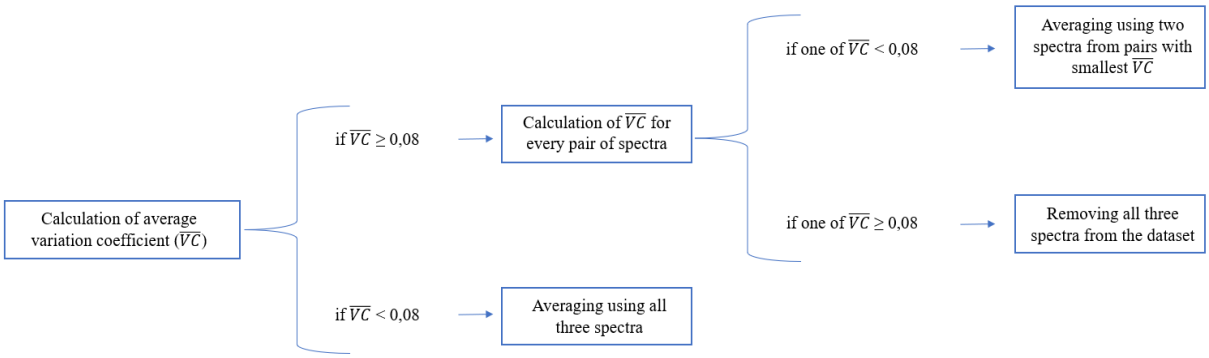


Figure 4.8. General scheme of filtering of outliers in measurements

4.2.4.3. Normalization of spectra

The anatomical site where spectra were taken, as well as the age, gender and phototype of the patients impact the shape and intensity of the measured spectral curves. Appropriate methods of spectral signal normalization have to be implemented since most of the data classification methods are sensitive to scale. In the present study, three normalization methods were compared: two maximum intensity (peak) normalization methods and a baseline standardization method.

The first peak normalization method consists in dividing each spectrum by its maximal intensity value lying in spectral range of [420:600] nm for the AF spectra (see equation 3) and in the range of [630:690] nm for the DR spectra (equation 4). The search range for finding the maximum for DR spectra was fixed so that the influence of hemoglobin absorption on the shape of the spectral curve is much lower than the influence of light scattering. The search range for normalization of the AF spectra was notably chosen to avoid the presence of the large amount of noise for all wavelengths above 600 nm.

$$AF_i^n(\lambda) = \frac{AF_i(\lambda)}{max[AF_i(420{-}600\ nm)]} \tag{4.3}$$

$$DR_i^n(\lambda) = \frac{DR_i(\lambda)}{max[DR_i(630{-}690\ nm)]} \tag{4.4}$$

Another normalization method, proposed by Tunnell et al. [33], consists in standardizing DR and AF spectra to a baseline or reference point as defined in equations 5 and 6.

$$AF_i^n(\lambda) = AF_i(\lambda) \cdot \frac{\overline{AF^H}}{max\left[\overline{AF_i^H}(\lambda)\right]} \tag{4.5}$$

$$DR_i^n(\lambda) = DR_i(\lambda) \cdot \frac{\overline{DR^H}}{max\left[\overline{DR_i^H}(630{-}690\ nm)\right]} \tag{4.6}$$

where $AF_i(\lambda)$ and $DR_i(\lambda)$ stand for the acquired AF and DR spectra of lesions or healthy skin measured near the lesion and *i* is the index of the lesion, $\overline{AF^H}$and $\overline{DR^H}$ relate to the average intensity over wavelength of the corresponding AF and DR spectra of healthy skin and $\overline{AF_i^H}(\lambda)$ and $\overline{DR_i^H}(\lambda)$ correspond to the average AF spectra of healthy skin measured near the corresponding lesion. Here, the main idea was to normalize the intensity of the spectra using spectra of healthy skin without changing the shape of the spectral curves since ratios $\frac{\overline{AF^H}}{\max\left[\overline{AF_i^H}(\lambda)\right]}$ and $\frac{\overline{DR^H}}{\max\left[\overline{DR_i^H}(\lambda)\right]}$ are constant scalar values. The 630-690 nm range of the DR-spectra was chosen by these authors as the wavelength interval where the contribution of light scattering to the shape of the spectral curve is mainly influenced by the hemoglobin and melanin absorption.

However, as mentioned above, anatomical site, phototype and other parameters of the skin have a significant influence not only on the peak intensity of the spectra, but also on their shape. As a consequence, the use of standardisation methods affecting only the scale of the spectra may not be sufficient. Therefore, we also implemented a normalization method with reference to the average shape of healthy spectra.

$$AF_i^n(\lambda) = AF_i(\lambda) \cdot \frac{\overline{AF^H}(\lambda)}{\overline{AF_i^H}(\lambda)} \tag{4.7}$$

$$DR_i^n(\lambda) = DR_i(\lambda) \cdot \frac{\overline{DR^H}(\lambda)}{\overline{DR_i^H}(\lambda)} \tag{4.8}$$

where $\overline{AF_i^H}(\lambda)$ and $\overline{DR_i^H}(\lambda)$ correspond to the average AF and DR spectra of healthy skin measured in the area around the respective lesion and *i* is the index of the lesion, $\overline{AF^H}(\lambda)$ and $\overline{DR^H}$relates to the average AF and DR spectra of healthy skin for each SDS. Thus, the main difference from the method proposed by Tunnell is that the normalisation ratio is not calculated on average, but for each wavelength separately.

4.2.4.4. Dimensionality reduction

The large number of points in each spectrum can make their analysis as a difficult task, since working with a large number of variables requires more computational resources, and can also negatively affect the accuracy of the classification algorithm. To avoid this, the last step of spectra preprocessing was the dimensionality reduction by feature extraction. Two methods were tested: principal component analysis (PCA) with different kernels and autoencoder (AE) with various layer architectures and activation functions. The choice of PCA and AE was due, on the one hand, to their high performance in similar studies and, on the other hand, to the large

number of customizable hyperparameters to improve the final classification accuracy. The selection of the optimal PCA-kernels, AE and activation functions, as well as of the corresponding parameters (number of PCs for PCA, number of layers and size of each layer for AE) was based on the final classification accuracy and the time required to apply the method.

In addition, according to the results obtained in Chapter 3, the *MinMax* Scaler was used for all extracted features.

4.2.4.5. Classification

Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and Linear Discriminant Analysis (LDA) were chosen as classifiers, as all these methods showed high efficiency in solving the classification problem in studies on skin cancer diagnosis and other pathologies using spectroscopy. The k-fold cross-validation method was used to avoid overtraining of the classification model due to incorrect division of the original dataset into training and test parts. The value of parameter k was chosen equal to 5 as it allowed us to obtain a test dataset of 40 or more samples for each iteration of the cross-validation. In all classification tests, the class sizes were equally balanced to avoid the influence of different number of samples per class on the final classification accuracy defined as:

$$Accuracy = \frac{\sum_1^N TP_i}{N} \tag{4.9}$$

where TP_i correspond to the number of correctly predicted samples for lesional classes (i = SCC, BCC, CI, CNI or AK) or healthy skin (i = H) and *N* relates to number of classes.

The samples to keep for the classes to be reduced were randomly selected before each series of calculations. The selection of the best performing classifier and corresponding hyperparameters was based on the average final classification accuracy for all cross-validation tests. A list of the methods used and the corresponding hyperparameters are presented in **Table 4.4**. The parameter variation ranges were chosen based on the results of a previous study [47]. The entire sequence starting from random sample selection was repeated five times to obtain statistically reliable results.

Table 4.4. Data analysis methods and optimized hyperparameters.

Name of the method	Optimized parameters
Dimensionality reduction	
Principal component analysis	kernel, number of components
Autoencoder	Number of layers, size of each layer, type of solver
Classification	
Support Vector Machine	Kernel, decision function, regularization parameter value
Multi-Layer Perceptron	Number of layers, size of each layer, type of solver
Linear Discriminant Analysis	-

Most of the results were obtained in binary classification mode in order to make a more detailed comparison with existing studies in this field. Some parts of the results for each of the two sets of classes were obtained in the multiclass classification mode, as this format is more suitable for application in the differential diagnosis of skin lesions.

4.2.5 Data Fusion

The last stage of spectra processing was the fusion of decisions for each sample using all the predictions obtained for each modality and each SDS. Fusion was performed using three different methods: Majority Voting (MV), voting with weighted voting for each modality and SDS (MO) and Stacking (S), which is the use of a machine learning classifier to make a diagnosis based on the predictions obtained after the classification step.

The grouping of the data for fusion was done at three different levels: fusion of all results, fusion of the results obtained for each modality and fusion of the results obtained for each SDS. Thus, the general scheme of data fusion is shown in **Figure 4.9**.

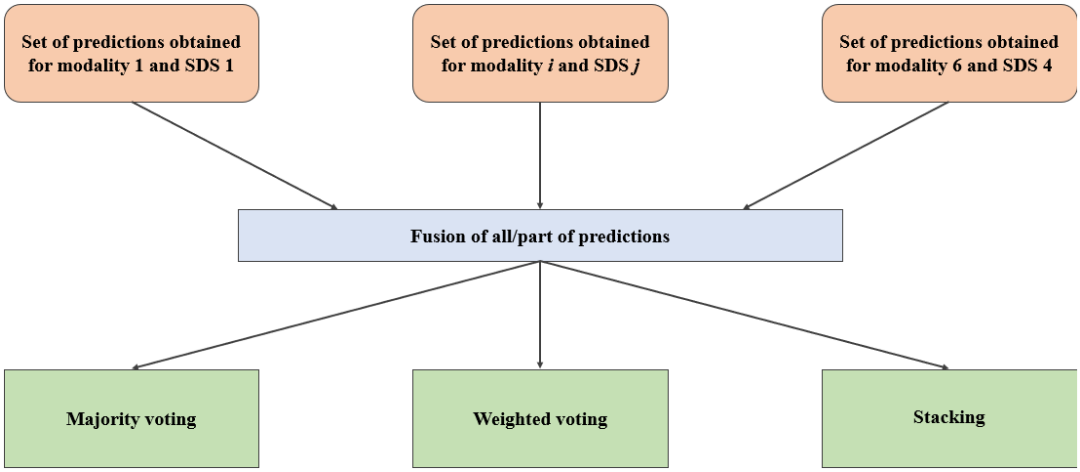


Figure 4.9. General scheme of fusion of decisions.

4.3 Results and Discussion

4.3.1. Selection of best methods and hyperparameters for the data processing pipeline

The selected values (or range of values) of the hyperparameters of each step of the spectroscopic data classification pipeline are provided in **Table 4.5**. They correspond to the highest accuracy values calculated after binary and multiclass classification for each combination of pairs of classes, modality and SDS separately. Since normalization has a significant effect on the performance of different data processing methods, it was also necessary to find, for each unique combination of modality, SDS and class pair, the normalization method that would provide the highest classification accuracy.

Table 4.5. Best values of hyperparameters retained for each method in the spectroscopic data classification pipeline.

Name of the method	List of hyperparameters	Hyperparameters value
Raw spectra filtering and correction		
Removing of Dark noise	-	-
Median filter	Kernel size	3
Savitzky–Golay filter	Size of the filtering window	~5 nm
	Polynomial order	4
Normalization		
No normalization	-	-
Intensity normalization	-	-
Shape normalization	-	-
Dimensionality reduction		
Principal component analysis	Number of PCs	15-20
Autoencoder	Number of hidden layers	3
	Size of the hidden layers	80, 40, 80
Classification		
Support vector machine	Kernel	Polynomial (2-3), linear
	Decision function shape	One vs Rest
Multi-layer Perceptron	Number of hidden layers	1
	Size of the layer	80-100
Linear Discriminant Analysis	-	-
Decision Fusion		
Majority voting	-	-
Stacking	Type of classifier	SVM, Random Forest with 80 trees
Weighted voting	-	-

4.3.2. Binary classification results

4.3.2.1 BCC, SCC, AK and H classes

The results of binary classification performed for the different pairs of BCC, SCC, AK and H classes, obtained for each combination of modality and SDS, are presented in **Figure 4.10**.

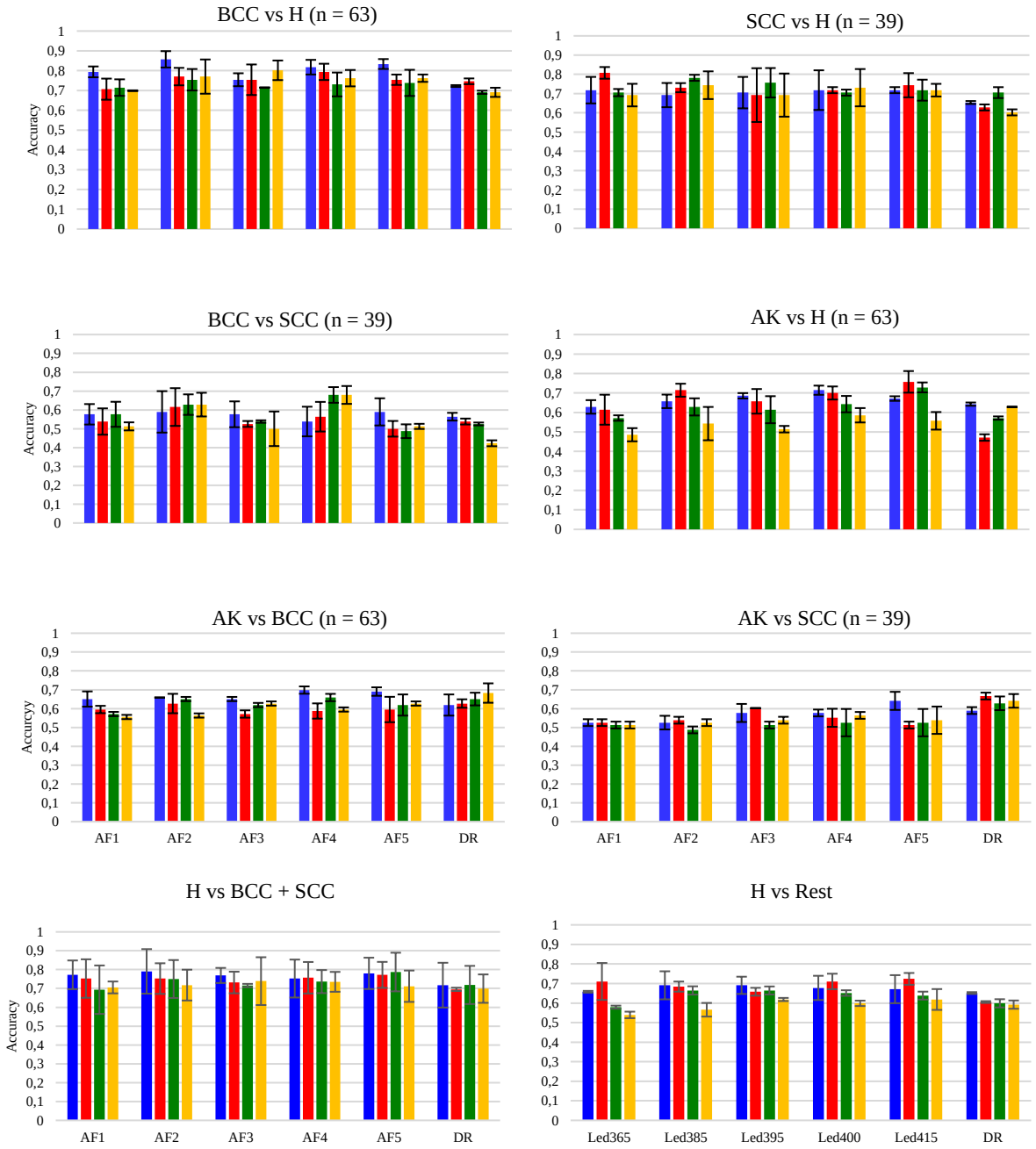


Figure 4.10 Results of pairwise binary classification for BCC, SCC, AK and H classes, for every modality (AF1..5, DR) and SDS (SDS_1 in blue, SDS_2 in red, SDS_3 in green and SDS_4 in orange).

All the results presented in **Figure 4.10** were obtained using shape normalization for AF spectra, no normalization for DR spectra and SVM classifier with linear kernel.

For H vs. BCC, the highest average accuracy values for single modalities separately are obtained for AF_2 (85,71%) and AF_5 (83,33%) at SDS_1. The maximum classification accuracy for H vs. AK is slightly lower (75.71% obtained for AF_5 at SDS_2) , which may be a consequence of the fact that the dataset contains samples for different stages of AK. For H vs. SCC, AF_1 at SDS_2 and AF_2 at SDS_3 provide the highest values with 80,77% and 78,21% respectively. For comparison between cancer classes BCC and SCC, best accuracy values were obtained for AF_4 modality at longest distances SDS_3 and SDS_4 (both 67,95%) Globally, the classification accuracy for pairs of classes containing AK is lower than 70%. For AK vs. Carcinomas, AF single modalities perform well again: the highest accuracy values obtained for AK vs BCC and AK vs SCC were 69,84% for AF_4 at SDS_1 and 64,10% for AF_5 at SDS_1 respectively. However, in case of AK vs SCC it can be observed that DR modality alone performs better than any other AF modality with 66,67% at SDS_2. These limited levels of accuracy obtained for AK vs. others can be related to the fact that the analyzed set of spectra contains spectra of skin AK samples of different histological subtypes “overlapping” in a single lesion (e.g. hypertrophic, atrophic, bowenoid, acantholytic, epidermolytic, lichenoid and pigmented), which spectroscopic features can be more or less close to healthy skin or to cancerous lesions.

The separation of samples into healthy and cancerous as well as healthy and pathological (cancerous plus AK) can also be of great interest in real diagnostic tasks. For H vs BCC+SCC, the highest classification accuracy was obtained for DR spectra at SDS_1 (76.83%) and AF_5 at SDS_3 (74.39%). For H vs Rest, the highest accuracy was obtained for AF_5 at SDS_2 (72.37%). Compared to the classification results for H vs BCC and H vs SCC, the classification accuracy for the H vs BCC + SCC case decreases by a few per cent, which can be explained by the larger size of the training and test dataset, as well as the difference of the more complex composition of the overall Cancerous class. In addition, as in the cases of AK vs others, classification accuracy decreases when AK class is added to the analysis (H vs Rest).

It is noticeable that, despite the close accuracy of the spectra classification results for different combinations of modality, SDS, normalization method and classifier, the best results for all pairs of classes were obtained using shape normalization and SVM classifier with linear kernel. However the efficiency of different combinations of data processing methods for classifying identical pairs of classes can be comparable or less than the standard deviations, which in most cases are 5-7%. This fact, in addition to the high level of diversity in the original

spectra, as well as the relatively small size of the data set, does not allow us to confidently conclude that this combination of classifier and normalization method will show the best results with other input data.

In addition, it is important to note that for each pair of classes, the most accurate results were obtained for a unique combination of modality and SDS, proving the greater efficiency of multimodal systems compared to unimodal systems.

The best accuracy results obtained after fusion of decisions for all pairs of classes are showed in Figure 6 and values presented in **Table 4.6**.

Table 4.6. Results of decision fusion for BCC, SCC, AK and H classes. Best values are in bold.

Classes (Number of samples for each class)	Majority voting, %	Weighted voting, %	Stacking, %
BCC vs H (63)	84,12	**86,51**	85,71
SCC vs H (39)	78,21	80,77	**83,33**
SCC vs BCC (39)	**67,95**	**67,95**	66,67
AK vs H (63)	66,15	**80,00**	**80,00**
AK vs BCC (63)	65,87	**70,63**	67,46
AK vs SCC (39)	65,38	**69,23**	**69,23**
H vs BCC + SCC	76,83	**81,71**	80,49
H vs Rest	69,74	**72,37**	**72,37**

The effectiveness of the application of decision fusion methods varies for different pairs of classes, Majority Voting performing the worst (just one bold value in Table 6). Using Weighted voting or Stacking , it was possible to increase the classification accuracy (compared to classification of spectra without the use of Data Fusion) from 3 up to 5% for pairs of classes H vs. others and from 1 up to 2% for pairs of classes containing only pathologies. These results confirm the interest of combining different AF and DR modalities together for improving the classification accuracy, which confirms the effectiveness of using multiple modalities instead of a single modality.

4.3.2.2 CI, CNI, AK and H classes

The results of binary classification performed for the different pairs of CI, CNI, AK and H classes, for each combination of modality and SDS, are presented in **Figure 4.11**.

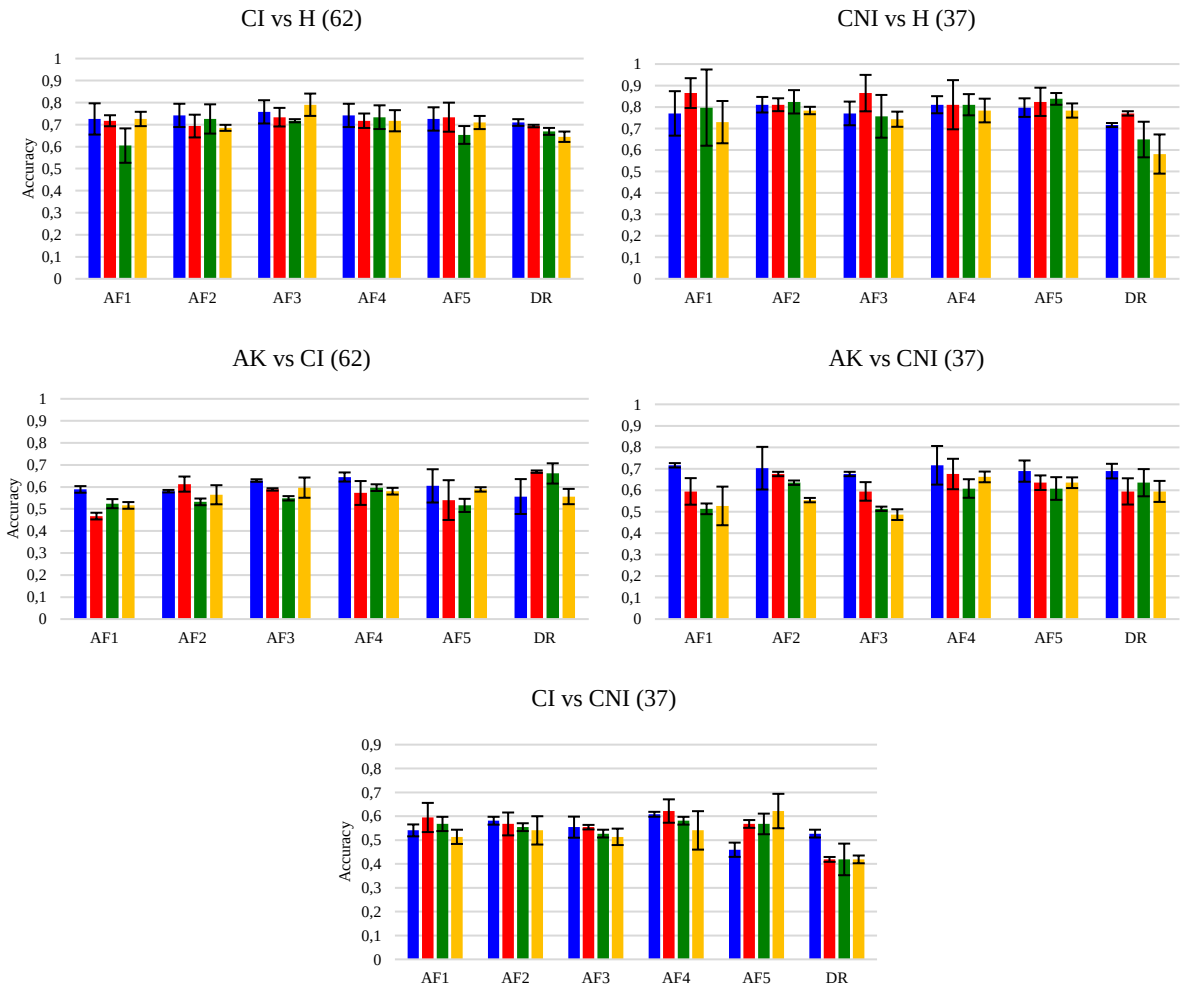


Figure 4.11. Results of classification for dividing cancer into invasive and non-invasive types for every modality (AF1..5, DR) and SDS (SDS_1 in blue, SDS_2 in red, SDS_3 in green and SDS_4 in orange).

The maximum accuracy values for pairs of classes CI vs H and CNI vs H were 79.03% and 86.48% obtained for AF_3 at SDS_4 and both AF_1 and AF_3 at SDS_2, respectively. At the same time, as in **Section 4.3.2.1**, the accuracy for pairs of classes containing AK is close to 70%: for AK vs CI, the maximum classification accuracy obtained for DR spectra measured at SDS_1 was 66.94%; for AK vs CNI, the highest accuracy value was obtained for AF_1 and AF_4 at SDS_1 and was 71.62%. The lowest classification accuracy among all pairs of classes (62.16%) was obtained for CI vs CNI, which may indicate a minor influence of cancer invasiveness on the characteristics of in vivo measured spectra.

As for the results shown in **Section 4.3.2.1**, the best results in all cases were achieved using shape normalization and SVM. However, in some cases the use of polynomial kernel led to better results than the use of linear kernel. However, it is important to notice that the

difference in classification accuracy values obtained for the same pairs of classes for different modalities and SDSs is comparable to the standard deviations. Considering this observation together with the fact that dataset has a small size, makes it impossible to clearly conclude that a particular modality or SDS provides better classification results than the others. It's also important to notice that the accuracy of discrimination for invasive cancer and healthy skin is much higher than that for non-invasive cancer and healthy skin, which may be due to the more significant changes in the optical properties of the skin associated with carcinogenesis. At the same time, the much smaller size of the dataset and high values of standard deviations may also be the reasons of the significant difference in classification accuracy for invasive and non-invasive skin carcinomas.

The next step was to use fusion of decisions by stacking for different groups of results obtained. The corresponding results are shown in Table 4.7.

Table 4.7. Results of decision fusion for CI, CNI, AK and H classes. Highest accuracies are in bold.

Classes (Number of samples for each class)	Majority voting	Weighted voting	Stacking
CI vs CNI (37)	60,81	**64,86**	56,76
CI vs H (62)	76,61	**79,03**	78,22
CNI vs H (37)	86,49	**87,83**	85,14
AK vs CI (62)	62,09	**66,93**	64,51
AK vs CNI (37)	70,27	**74,32**	71,62

As in **Section 4.3.2.1**, the use of decision fusion improved classification accuracy for all pairs of classes from 2 to 3%.

4.3.3. Multiclass classification results

The use of a binary classification may not be sufficient for clinical applications.. Taking this fact into account,, we also performed a series of multiclass tests whose, results are presented in **Figure 4.12**.

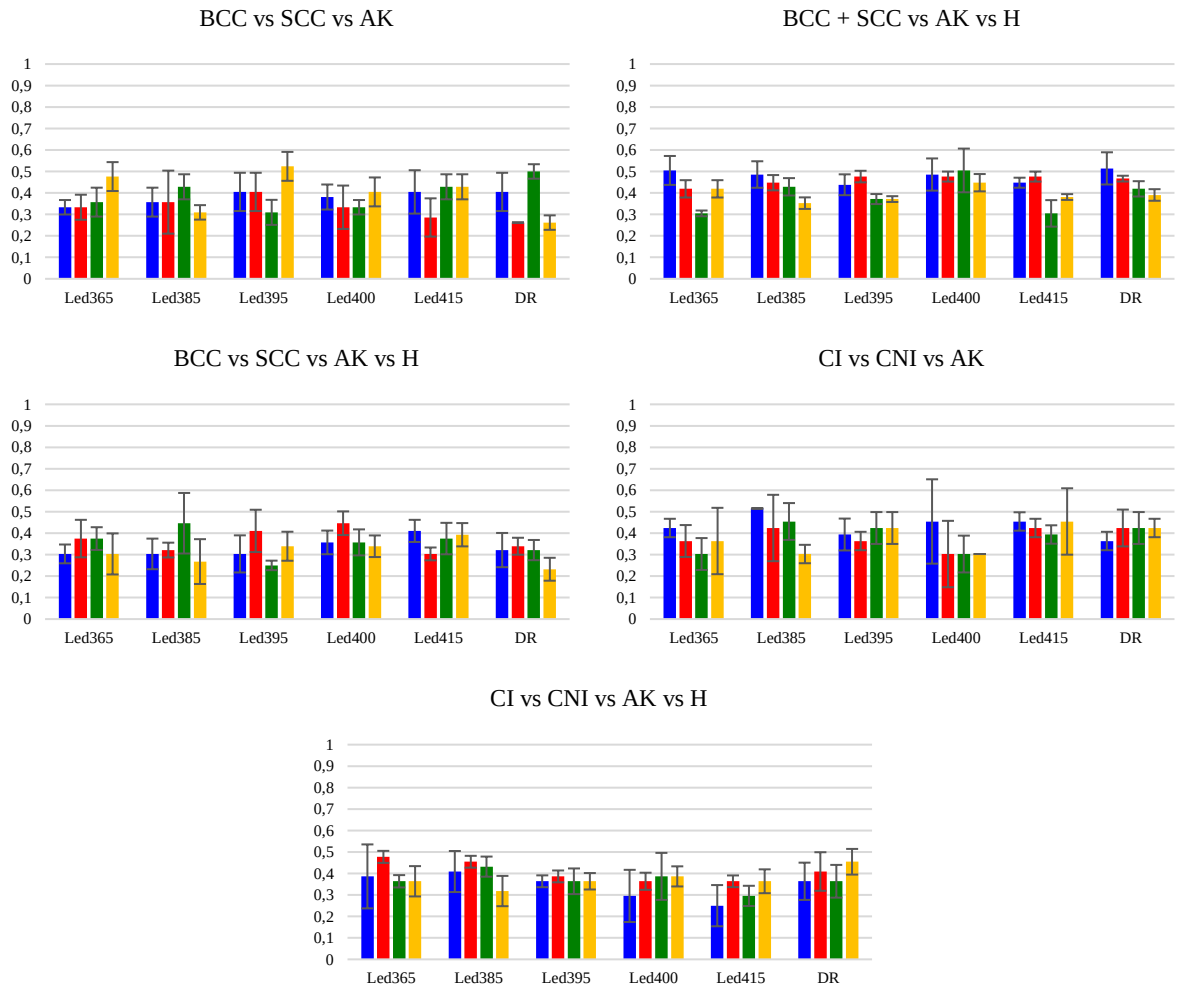


Figure 4.12. Results of multiclass classifications for every modality (AF1..5, DR) and SDS (SDS_1 in blue, SDS_2 in red, SDS_3 in green and SDS_4 in orange).

When spectra were divided into more than two classes, the classification accuracy did not exceed 60% in any of the cases: for BCC vs SCC vs AK the maximum classification accuracy obtained was 52.38%, for BCC + SCC vs AK vs H the best accuracy was 51.43% and for BCC vs SCC vs AK vs H only a 44.64% accuracy was reached. By sorting all cancer samples into invasive or non-invasive types, the maximum classification accuracy obtained was 51.51% for CI vs CNI vs AK and 47.72% for CI vs CNI vs AK vs H, respectively. Thus, the obtained accuracy for all multiclass classification attempts is significantly lower than that of binary classification. The use of decision fusion whose results are presented in **Table 4.8**, also failed to achieve accuracy values for skin cancer or AK that exceeded 60% for at least one of the class combinations studied.

Table 4.8. Best results of single modality multiclass classification by main diagnostic classes. Best values in bold.

Classes (Number of samples for each class)	Majority voting	Weighted voting	Stacking
SCC vs BCC vs AK (46)	40,47	**57,14**	40,47
SCC + BCC vs H vs AK (84)	49,53	**56,19**	53,33
SCC vs BCC vs H vs AK (46)	42,86	**50,00**	39,28
CI vs CNI vs AK (37)	51,52	**57,58**	39,40
CI vs CNI vs AK vs H (37)	47,27	**56,82**	47,27

4.3.4. Evaluation of the results obtained

4.3.4.1. In relation to the preclinical study

A direct comparison of the results of the preclinical and clinical studies may not be completely correct, since in the preclinical study the measurements were performed essentially on the non pigmented hairless skin of mice back with benign and potentially precancerous lesions, while the clinical study dealt with human skin samples of various phototype and anatomical sites with precancerous lesions (AK) or/and carcinomas. At the same time, the list methods of spectra processing is almost identical in both studies, and mouse skin models are well studied in the context of NMSC carcinogenesis, so the comparison of the results of clinical and preclinical studies is of some interest.

The maximum multiclass classification accuracy of precancerous skin lesions in the preclinical study (94.31%) was significantly higher than all classification accuracy values obtained in the clinical study, including even the binary classification results (no higher than 87% mostly below 75%). Such a large difference in the classification accuracy of skin lesions in favour of the preclinical study can be explained by the smaller skin thickness of the mice: the total thickness of the epidermis and dermis on the body of mice is about 300 μm, while for humans the similar value can reach several millimetres. According to [240, 241], the penetration depth of detected photons in the SpectroLive Device can reach two millimetres, which is more than enough for complete penetration of radiation into the skin of mice. Thus, the resulting spectrum obtained from the skin of mice contains information about changes in the properties of all skin layers, which has a positive effect on the diagnostic potential of the method.

Another important factor is the difference in the level of variability in skin features an properties specific to the dataset being analysed. For example, a group of 20 laboratory mice taken from the same population for a preclinical study (no hair, the same anatomical site, no pigmentation, same age, same gender, same carcinogenesis induction conditions and same

living conditions) should have much less inter- and intra-individual variations in both healthy and pathological skin features than a group of 131 patients with different gender, age, phototype, lesion localisation and background, which includes different exposure of the body to UV radiation, bad habits etc..

At the same time, it is important to note that the separation of lesions and healthy tissue in both studies is performed quite well, which confirms the promising experimental technique and the effectiveness of the chosen methodology for spectra analysis.

4.3.4.2. In relation to other studies on the use of AF and DR spectra for skin cancer diagnosis

In addition to comparing the results of the preclinical and clinical studies conducted at CRAN, it is also of great interest to compare the results obtained with the results of clinical studies on the use of AF and DR spectroscopy performed by other teams.

The accuracy values obtained after applying decision fusion were close to or over 80% for all binary classification tests that included cancerous and healthy samples, which is close to both existing multimodal studies [124, 126-129, 130-133] and single-modality studies [62, 64, 75-82] in the field of in vivo skin cancer diagnosis. It has to be noticed that the size of the dataset was quite small (less than 30 samples per class) in seven [20, 119–122, 147, 164] of fifteen single modality studies and in three [33, 34, 175] of eight multimodal studies on the diagnosis of non-melanoma skin cancer listed in the Chapter 1, which contained diagnostic accuracy metrics. Thus, the diagnostic accuracy values of 80% (or higher) obtained in these references may be slightly overestimated.

In addition, in contrast to most of other studies, several different methods were tested at most stages of the data processing pipeline, thus reducing to a quite low level the chance that, due to an incorrectly chosen preprocessing, classification or data fusion methods, the classification accuracy significantly decreases.

4.3.5. The use of other data pre-processing and processing methods

The list of data pre-processing and processing methods used in the preclinical and then clinical studies was developed based on existing research on the use of optical biopsy techniques for skin cancer diagnosis. This approach allowed us to identify the most common methods that are likely to be the most effective: PCA for dimensionality reduction by feature extraction, SVM/MLP/LDA for classification, voting for fusion of decisions etc. It also facilitated the comparison of our results with the results those obtained other studies in this

field. However, the number of such studies is not very large, which greatly limits the number of possible methods to be implemented.

The limited size of the list of methods may not be important if the use of methods from this list provides good performance of data analysis, which in the case of diagnostic studies usually means high accuracy in identifying types or parameters of pathologies. However, if the accuracy values obtained with previously tested methods are lower than expected, it may be worthwhile to test methods that have not been used in similar studies.

The classification accuracy of the mouse skin spectra obtained in the preclinical study (Chapter 3) exceeded 90% in multiclass mode, and hence no testing of other preprocessing or data processing methods was required. In contrast, the classification accuracy of human skin spectra obtained in the clinical study (Chapter 4) did not exceed 60% in multiclass mode for all combinations of classes tested. In order to improve accuracy, several other methods and strategies were tested at different stages of the processing pipeline for the data obtained in the clinical trial.

At the normalisation stage, ratios of intensities for different wavelengths were tested as normalisation coefficients, with these coefficients used for both multiplication by spectra and division of spectra by spectra. However, there was no improvement in the overall classification accuracy of the spectra.

In the dimensionality reduction stage, Feature selection methods such as Variance Threshold related to Filter methods and feature selection based on spectra classification accuracy related to Wrapper methods were tested. The use of Variance Threshold only degraded the overall classification accuracy for all class combinations relative to the data pipelines where PCA was used for dimensionality reduction. In contrast, when feature selection was based on final classification accuracy, the results obtained were comparable to those obtained using PCA. However, the use of this method required much more computational resources and, as a consequence, more time consuming. In addition, despite the fact that there was no transition to a new feature space, the selected features were not interpreted in any way.

In addition to standard feature selection/extraction methods, it was also decided to test spectroscopic features for classification, such as area under the curve, positions and intensity ratios of fluorescence (for AF spectra) and absorption (for DR spectra) peaks, interpolation coefficients of the spectral curve or its sections, and others. Unfortunately, the use of these methods also did not improve the accuracy of spectra classification.

At the classification stage, it was decided not to add new classification methods, as all the methods tested in this work have proven to be highly effective in many other studies. Instead, we tested different options for sorting the obtained spectra into classes and subclasses, combining spectra with clinical parameters (age of patients and their skin age, skin phototype, lesion localisation) as they have a strong influence on the properties of the skin [242]. Stepwise classification was also tested instead of multiclass classification. However, none of these approaches improved the accuracy of spectra classification. At the same time, it is important to note that using multiple clinical parameters to sort spectra may make sense, however, this is not possible at this stage of the study due to the small size of the dataset.

Thus, despite the many combinations of methods and approaches tested, there was no improvement within the spectra classification accuracy compared to the results presented earlier in this chapter.

4.4. Conclusion

This chapter presents the results of the classification of AF and DR spectra measured in vivo on human skin with non-melanoma skin cancers and actinic keratosis as part of a clinical study that included 131 patients.

The data processing pipeline developed included the stages of noise removal and filtering, detection of unstable measurements, standardisation/normalisation, dimensionality reduction by feature extraction, classification and data fusion. The candidate methods in each stage were selected based on the effectiveness of their application in previous studies by our team or in other studies on a similar topic.

The selection of the final combination of methods was based on the classification accuracy of the spectra obtained after implementation of data fusion. However, the highest classification accuracy for all pairs of classes was obtained using shape normalization by corresponding average healthy skin spectra, PCA for feature extraction, SVM with linear kernel as a classifier and the decision fusion by weighted voting as a data fusion method.

Spectra classification was performed in binary and multiclass modes for different pairs and combinations of classes. It is also important to note that two types of labels were used for classification of spectra corresponding to skin carcinoma samples: those related to the diagnostic type of skin carcinoma (BCC or SCC) or to their invasiveness (invasive cancer and non-invasive cancer).

The highest accuracy values among all pairs of classes were obtained for Healthy skin vs Cancerous lesions and exceeded 80%. When Actinic Keratosis was compared with carcinomas or healthy skin, the classification accuracy decreased and was 60-70%. At the same time, the multiclass classification accuracy for all tested class combinations did not exceed 60%. It is important to notice that the use of Data Fusion for all modalities and SDS increased the classification accuracy of spectra in all cases from 3 up to 5%.

At the end of this chapter, the results obtained are compared with the results of a preclinical study on mouse skin and the results of clinical studies on the use of AF and DR spectroscopy for the diagnosis of NMSC conducted in other laboratories.

General Conclusion and Prospects

General Conclusion

The main objective of this thesis was to combine machine learning methods and multimodal spectroscopic data acquired in vivo on skin with non-melanoma skin cancer and precancerous lesions in order to validate efficient processing pipeline optimizing the automatic discrimination between multiple classes of benign and malignant lesions hence optimizing the diagnostic accuracy.

The first step was to review the literature on the use of optical techniques for the diagnosis of non-melanoma skin cancer and identify a list of spectra processing methods, including machine learning techniques, that were used in these studies (Chapters 1 and 2).

The next step (Chapter 3) was to develop and quantify the performance of selected methods on a set of spectra obtained in vivo in a preclinical study of mice with benign and precancerous skin lesions. SVM, MLP and LDA classifiers in combination with PCA for Feature extraction and Data Fusion by Stacking or Weighted Voting were found to provide pathology recognition accuracy of over 90% in multiclass mode and over 97% in binary mode.

The final step (Chapter 4) was to adapt and optimize the resulting data pipeline on spectra acquired in vivo as part of a clinical skin study of patients with invasive and non-invasive NMSC and AK. The final accuracy in recognizing skin pathologies in multiclass mode was from 65 to 75% in most cases, and in some cases did not exceed 60%. This confirms that the skin structure of humans is more complex compared to mice, as well as the fact that the level of variability of skin parameters among the participating patients is also much higher than in laboratory mice. In addition, the results suggest the need for continued clinical trials as well as further development of the study concept.

However, it is also important to note that in binary classification mode the accuracy of discrimination between spectra measured on carcinoma and healthy skin samples exceeded

80% in all cases, which proves the high diagnostic potential of the combination of experimental and data processing methods proposed in the presented study.

Prospects

There are several directions of development for the current study, the first of which is to increase the size of the dataset by adding spectra measured on new patients. Adding new spectra to the dataset may help to obtain more reliable machine learning models, also for multiclass tests.

Another goal is to find optimal parameters for normalizing or sorting the obtained spectra in order to improve the accuracy of the obtained machine learning models. However, the solution of this problem is almost impossible without increasing the size of the dataset.

The next step after finding and optimizing the list of parameters for normalizing or sorting spectra should be to improve the interpretability of the results by adding information about changes in the optical properties of the skin that may be directly associated with carcinogenesis. Obtaining such information is possible by using Monte Carlo simulations or other methods to solve the Radiative Transfer Equation. Application of these methods at this stage of the study is possible , but does not make much sense, as the level of variability in the form of spectra belonging to the same class is too high.

Once all the above steps have been completed, there is a wider application of the SpectroLive device in clinical practice to assist in the diagnosis of non-melanoma skin cancer and related precancerous lesions.

List of references

1. Bray F, Laversanne M, Sung H, et al (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74:229–263. https://doi.org/10.3322/caac.21834
2. Khayyati Kohnehshahri M, Sarkesh A, Mohamed Khosroshahi L, et al (2023) Current status of skin cancers with a focus on immunology and immunotherapy. Cancer Cell Int 23:174. https://doi.org/10.1186/s12935-023-03012-7
3. Roser M, Ritchie H (2024) Cancer. Our World Data
4. Ciążyńska M, Kamińska-Winciorek G, Lange D, et al (2021) The incidence and clinical analysis of non-melanoma skin cancer. Sci Rep 11:4337. https://doi.org/10.1038/s41598-021-83502-8
5. Eisemann N, Waldmann A, Geller AC, et al (2014) Non-Melanoma Skin Cancer Incidence and Impact of Skin Cancer Screening on Incidence. J Invest Dermatol 134:43–50. https://doi.org/10.1038/jid.2013.304
6. Mogensen M, Jemec GBE (2007) Diagnosis of Nonmelanoma Skin Cancer/Keratinocyte Carcinoma: A Review of Diagnostic Accuracy of Nonmelanoma Skin Cancer Diagnostic Tests and Technologies. Dermatol Surg 33:1158
7. Nault A, Zhang C, Kim K, et al (2015) Biopsy Use in Skin Cancer Diagnosis: Comparing Dermatology Physicians and Advanced Practice Professionals. JAMA Dermatol 151:899–902. https://doi.org/10.1001/jamadermatol.2015.0173
8. Skaggs R, Coldiron B (2021) Skin biopsy and skin cancer treatment use in the Medicare population, 1993 to 2016. J Am Acad Dermatol 84:53–59. https://doi.org/10.1016/j.jaad.2020.06.030
9. Narayanamurthy V, Padmapriya P, Noorasafrin A, et al (2018) Skin cancer detection using non-invasive techniques. RSC Adv 8:28095–28130. https://doi.org/10.1039/C8RA04164D
10. Calin MA, Parasca SV, Savastru R, et al (2013) Optical techniques for the noninvasive diagnosis of skin cancer. J Cancer Res Clin Oncol 139:1083–1104. https://doi.org/10.1007/s00432-013-1423-3

11. Kim JA, Wales DJ, Yang G-Z (2020) Optical spectroscopy for in vivo medical diagnosis—a review of the state of the art and future perspectives. Prog Biomed Eng 2:042001. https://doi.org/10.1088/2516-1091/abaaa3

12. Rey-Barroso L, Peña-Gutiérrez S, Yáñez C, et al (2021) Optical Technologies for the Improvement of Skin Cancer Diagnosis: A Review. Sensors 21:252. https://doi.org/10.3390/s21010252

13. Garcia-Uribe A, Zou J, Duvic M, et al (2012) In Vivo Diagnosis of Melanoma and Nonmelanoma Skin Cancer Using Oblique Incidence Diffuse Reflectance Spectrometry. Cancer Res 72:2738–2745. https://doi.org/10.1158/0008-5472.CAN-11-4027

14. Zhang Y, Moy AJ, Feng X, et al (2019) Physiological model using diffuse reflectance spectroscopy for nonmelanoma skin cancer diagnosis. J Biophotonics 12:e201900154. https://doi.org/10.1002/jbio.201900154

15. Upile T, Jerjes W, Radhi H, et al (2012) Elastic scattering spectroscopy in assessing skin lesions: An "*in vivo*" study. Photodiagnosis Photodyn Ther 9:132–141. https://doi.org/10.1016/j.pdpdt.2011.12.003

16. Panjehpour M, Julius CE, Phan MN, et al (2002) Laser-induced fluorescence spectroscopy for in vivo diagnosis of non-melanoma skin cancers. Lasers Surg Med 31:367–373. https://doi.org/10.1002/lsm.10125

17. Andreeva V, Aksamentova E, Muhachev A, et al (2022) Preoperative AI-Driven Fluorescence Diagnosis of Non-Melanoma Skin Cancer. Diagnostics 12:72. https://doi.org/10.3390/diagnostics12010072

18. Maciel VH, Correr WR, Kurachi C, et al (2017) Fluorescence spectroscopy as a tool to *in vivo* discrimination of distinctive skin disorders. Photodiagnosis Photodyn Ther 19:45–50. https://doi.org/10.1016/j.pdpdt.2017.03.014

19. Ouyang Y, Liu Y, Wang ZM, et al (2021) FLIM as a Promising Tool for Cancer Diagnosis and Treatment Monitoring. Nano-Micro Lett 13:133. https://doi.org/10.1007/s40820-021-00653-z

20. Ferreira Lima AM, Daniel CR, Navarro RS, et al (2019) Discrimination of non-melanoma skin cancer and keratosis from normal skin tissue *in vivo* and *ex vivo* by Raman spectroscopy. Vib Spectrosc 100:131–141. https://doi.org/10.1016/j.vibspec.2018.11.009

21. Zhang J, Fan Y, Song Y, Xu J (2018) Accuracy of Raman spectroscopy for differentiating skin cancer from normal tissue. Medicine (Baltimore) 97:e12022. https://doi.org/10.1097/MD.0000000000012022

22. Zhao J, Zeng H, Kalia S, Lui H (2019) Incorporating patient demographics into Raman spectroscopy algorithm improves in vivo skin cancer diagnostic specificity. Transl Biophotonics 1:e201900016. https://doi.org/10.1002/tbio.201900016

23. Feng X, Fox MC, Reichenberg JS, et al (2019) Biophysical basis of skin cancer margin assessment using Raman spectroscopy. Biomed Opt Express 10:104–118. https://doi.org/10.1364/BOE.10.000104

24. Delrue C, Speeckaert R, Oyaert M, et al (2023) From Vibrations to Visions: Raman Spectroscopy's Impact on Skin Cancer Diagnostics. J Clin Med 12:7428. https://doi.org/10.3390/jcm12237428

25. Ilișanu M-A, Moldoveanu F, Moldoveanu A (2023) Multispectral Imaging for Skin Diseases Assessment—State of the Art and Perspectives. Sensors 23:3888. https://doi.org/10.3390/s23083888

26. Petracchi B, Torti E, Marenzi E, Leporati F (2024) Acceleration of Hyperspectral Skin Cancer Image Classification through Parallel Machine-Learning Methods. Sensors 24:1399. https://doi.org/10.3390/s24051399

27. Huang H-Y, Hsiao Y-P, Karmakar R, et al (2023) A Review of Recent Advances in Computer-Aided Detection Methods Using Hyperspectral Imaging Engineering to Detect Skin Cancer. Cancers 15:5634. https://doi.org/10.3390/cancers15235634

28. Filoni A, Alaibac M (2019) Reflectance Confocal Microscopy in Evaluating Skin Cancer: A Clinicians's Perspective. Front Oncol 9:. https://doi.org/10.3389/fonc.2019.01457

29. Olsen J, Holmes J, Jemec GBE (2018) Advances in optical coherence tomography in dermatology—a review. J Biomed Opt 23:040901. https://doi.org/10.1117/1.JBO.23.4.040901

30. Xiong Y-Q, Mo Y, Wen Y-Q, et al (2018) Optical coherence tomography for the diagnosis of malignant skin tumors: a meta-analysis. J Biomed Opt 23:020902. https://doi.org/10.1117/1.JBO.23.2.020902

31. Wan B, Ganier C, Du-Harpur X, et al (2021) Applications and future directions for optical coherence tomography in dermatology*. Br J Dermatol 184:1014–1022. https://doi.org/10.1111/bjd.19553

32. Tuchin VV, Popp J, Zakharov V (2020) Multimodal Optical Diagnostics of Cancer. Springer International Publishing, Cham

33. Lim L, Nichols BS, Migden MR, et al (2014) Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis. J Biomed Opt 19:117003. https://doi.org/10.1117/1.JBO.19.11.117003

34. Thompson AJ, Coda S, Sørensen MB, et al (2012) In vivo measurements of diffuse reflectance and time-resolved autofluorescence emission spectra of basal cell carcinomas. J Biophotonics 5:240–254. https://doi.org/10.1002/jbio.201100126

35. Borisova E, Pavlova E, Kundurjiev T, et al (2014) Light-induced autofluorescence and diffuse reflectance spectroscopy in clinical diagnosis of skin cancer. In: Biophotonics: Photonic Solutions for Better Health Care IV. SPIE, pp 251–258

36. Carpenter DJ, Sajisevi MB, Chapurin N, et al (2018) Noninvasive optical spectroscopy for identification of non-melanoma skin cancer: Pilot study. Lasers Surg Med 50:246–252. https://doi.org/10.1002/lsm.22786

37. Bratchenko IA, Artemyev DN, Myakinin OO, et al (2017) Combined Raman and autofluorescence ex vivo diagnostics of skin cancer in near-infrared and visible regions. J Biomed Opt 22:027005. https://doi.org/10.1117/1.JBO.22.2.027005

38. Feng X, Moy AJ, Nguyen HTM, et al (2017) Raman active components of skin cancer. Biomed Opt Express 8:2835–2850. https://doi.org/10.1364/BOE.8.002835

39. Fedorov Kukk A, Wu D, Gaffal E, et al (2022) Multimodal system for optical biopsy of melanoma with integrated ultrasound, optical coherence tomography and Raman spectroscopy. J Biophotonics 15:e202200129. https://doi.org/10.1002/jbio.202200129

40. Blondel W, Delconte A, Khairallah G, et al (2021) Spatially-Resolved Multiply-Excited Autofluorescence and Diffuse Reflectance Spectroscopy: SpectroLive Medical Device for Skin In Vivo Optical Biopsy. Electronics 10:243. https://doi.org/10.3390/electronics10030243

41. AMOUROUX M, Blondel W, DELCONTE A (2021) Medical device for fibred bimodal optical spectroscopy

42. Amouroux M, Díaz-Ayil G, Blondel WCPM, et al (2009) Classification of ultraviolet irradiated mouse skin histological stages by bimodal spectroscopy: multiple excitation autofluorescence and diffuse reflectance. J Biomed Opt 14:014011. https://doi.org/10.1117/1.3077194

43. Díaz-Ayil G, Amouroux M, Blondel WCPM, et al (2009) Bimodal spectroscopic evaluation of ultra violet-irradiated mouse skin inflammatory and precancerous stages: instrumentation, spectral feature extraction/selection and classification (k-NN, LDA and SVM). Eur Phys J - Appl Phys 47:12707. https://doi.org/10.1051/epjap/2009032

44. Elsen T, Fauvel C, Khairallah G, et al (2024) A dataset of optical spectra and clinical features acquired on human healthy skin and on skin carcinomas. Data Brief 53:110163. https://doi.org/10.1016/j.dib.2024.110163

45. Lin WM, Yuan X, Yuen PW, et al (2003) Combined classifier for discriminating cancerous tissue from normal tissue using light-induced autofluorescence. In: Diagnostic Optical Spectroscopy in Biomedicine II (2003), paper 5141_177. Optica Publishing Group, p 5141_177

46. Abdat F, Amouroux M, Guermeur Y, Blondel W (2012) Hybrid feature selection and SVM-based classification for mouse skin precancerous stages diagnosis from bimodal spectroscopy. Opt Express 20:228–244. https://doi.org/10.1364/OE.20.000228

47. Kupriyanov V, Blondel W, Daul C, et al (2023) Implementation of data fusion to increase the efficiency of classification of precancerous skin states using in vivo bimodal spectroscopic technique. J Biophotonics 16:e202300035. https://doi.org/10.1002/jbio.202300035

48. Benson HAE (2011) Skin Structure, Function, and Permeation. In: Topical and Transdermal Drug Delivery. John Wiley & Sons, Ltd, pp 1–22

49. (2024) Human skin | Definition, Layers, Types, & Facts | Britannica. https://www.britannica.com/science/human-skin. Accessed 28 May 2024

50. Owda A, Salmon N, Casson A, Owda M (2020) The Reflectance of Human Skin in the Millimeter-Wave Band. Sensors 20:1480. https://doi.org/10.3390/s20051480

51. Baroni A, Buommino E, De Gregorio V, et al (2012) Structure and function of the epidermis related to barrier properties. Clin Dermatol 30:257–262. https://doi.org/10.1016/j.clindermatol.2011.08.007

52. Epidermis (Outer Layer of Skin): Layers, Function, Structure. In: Clevel. Clin. https://my.clevelandclinic.org/health/body/21901-epidermis. Accessed 28 May 2024

53. Marks JG, Miller JJ (2017) Lookingbill and Marks' Principles of Dermatology. Elsevier Health Sciences

54. Lopez-Ojeda W, Pandey A, Alhajj M, Oakley AM (2024) Anatomy, Skin (Integument). In: StatPearls. StatPearls Publishing, Treasure Island (FL)

55. Prost-Squarcioni C, Fraitag S, Heller M, Boehm N (2008) [Functional histology of dermis]. Ann Dermatol Venereol 135:1S5-20. https://doi.org/10.1016/S0151-9638(08)70206-0

56. Dermis (Middle Layer of Skin): Layers, Function & Structure. In: Clevel. Clin. https://my.clevelandclinic.org/health/body/22357-dermis. Accessed 29 May 2024

57. Reticular layer of dermis. In: Kenhub. https://www.kenhub.com/en/library/anatomy/reticular-layer-of-dermis. Accessed 29 May 2024

58. Raskin RE, Meyer D (2015) Canine and Feline Cytology - E-Book: Canine and Feline Cytology - E-Book. Elsevier Health Sciences

59. Homepage – IARC. https://www.iarc.who.int. Accessed 16 Jul 2024

60. Samarasinghe V, Madan V (2012) Nonmelanoma Skin Cancer. J Cutan Aesthetic Surg 5:3–10. https://doi.org/10.4103/0974-2077.94323

61. Linares MA, Zakaria A, Nizran P (2015) Skin Cancer. Prim Care 42:645–659. https://doi.org/10.1016/j.pop.2015.07.006

62. Rubin Adam I., Chen Elbert H., Ratner Désirée (2005) Basal-Cell Carcinoma. N Engl J Med 353:2262–2269. https://doi.org/10.1056/NEJMra044151

63. Dinehart SM, Dodge R, Stanley WE, et al (1992) Basal Cell Carcinoma Treated with Mohs Surgery: A Comparison of 54 Younger Patients with 1050 Older Patients. Dermatol Surg 18:560. https://doi.org/10.1111/j.1524-4725.1992.tb03509.x

64. Wong CSM, Strange RC, Lear JT (2003) Basal cell carcinoma. BMJ 327:794–798

65. Choi Y-H, Tak Y-S, Rho S, Hwang E (2013) Skin feature extraction and processing model for statistical skin age estimation. Multimed Tools Appl 64:227–247. https://doi.org/10.1007/s11042-011-0987-7

66. Cives M, Mannavola F, Lospalluti L, et al (2020) Non-Melanoma Skin Cancers: Biological and Clinical Features. Int J Mol Sci 21:5394. https://doi.org/10.3390/ijms21155394

67. Niculet E, Craescu M, Rebegea L, et al (2022) Basal cell carcinoma: Comprehensive clinical and histopathological aspects, novel imaging tools and therapeutic approaches (Review). Exp Ther Med 23:60. https://doi.org/10.3892/etm.2021.10982

68. Ramachandran S, Fryer AA, Smith AG, et al (2000) Basal cell carcinoma. Cancer 89:1012–1018. https://doi.org/10.1002/1097-0142(20000901)89:5<1012::AID-CNCR10>3.0.CO;2-O

69. Marcil I, Stern RS (2000) Risk of Developing a Subsequent Nonmelanoma Skin Cancer in Patients With a History of Nonmelanoma Skin Cancer: A Critical Review of the Literature and Meta-analysis. Arch Dermatol 136:1524–1530. https://doi.org/10.1001/archderm.136.12.1524

70. Webb JL, Burns RE, Brown HM, et al (2009) Squamous cell carcinoma. Compend Contin Educ Vet 31:E9

71. Alam Murad, Ratner Désirée (2001) Cutaneous Squamous-Cell Carcinoma. N Engl J Med 344:975–983. https://doi.org/10.1056/NEJM200103293441306

72. Kallini JR, Hamed N, Khachemoune A (2015) Squamous cell carcinoma of the skin: epidemiology, classification, management, and novel trends. Int J Dermatol 54:130–140. https://doi.org/10.1111/ijd.12553

73. Shanmugaratnam K (2012) Histological Typing of Tumours of the Upper Respiratory Tract and Ear. Springer Science & Business Media

74. Brantsch KD, Meisner C, Schönfisch B, et al (2008) Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. Lancet Oncol 9:713–720. https://doi.org/10.1016/S1470-2045(08)70178-5

75. Callen JP, Bickers DR, Moy RL (1997) Actinic keratoses. J Am Acad Dermatol 36:650–653. https://doi.org/10.1016/S0190-9622(97)70265-2

76. Olsen EA, Lisa Abernethy M, Kulp-Shorten C, et al (1991) A double-blind, vehicle-controlled study evaluating masoprocol cream in the treatment of actinic keratoses on the head and neck. J Am Acad Dermatol 24:738–743. https://doi.org/10.1016/0190-9622(91)70113-G

77. Schleusener J, Gluszczynska P, Reble C, et al (2015) In vivo study for the discrimination of cancerous and normal skin using fibre probe-based Raman spectroscopy. Exp Dermatol 24:767–772. https://doi.org/10.1111/exd.12768

78. Reinehr CPH, Bakos RM (2020) Actinic keratoses: review of clinical, dermoscopic, and therapeutic aspects'. An Bras Dermatol 94:637–657. https://doi.org/10.1016/j.abd.2019.10.004

79. Röwert-Huber J, Patel MJ, Forschner T, et al (2007) Actinic keratosis is an early in situ squamous cell carcinoma: a proposal for reclassification. Br J Dermatol 156:8–12. https://doi.org/10.1111/j.1365-2133.2007.07860.x

80. Ahmady S, Jansen MHE, Nelemans PJ, et al (2022) Risk of Invasive Cutaneous Squamous Cell Carcinoma After Different Treatments for Actinic Keratosis: A Secondary Analysis of a Randomized Clinical Trial. JAMA Dermatol 158:634–640. https://doi.org/10.1001/jamadermatol.2022.1034

81. Youl PH, Baade PD, Janda M, et al (2007) Diagnosing skin cancer in primary care: how do mainstream general practitioners compare with primary care skin cancer clinic doctors? Med J Aust 187:215–220. https://doi.org/10.5694/j.1326-5377.2007.tb01202.x

82. van der Geer S, Kleingeld PAM, Snijders CCP, et al (2015) Development of a Non-Melanoma Skin Cancer Detection Model. Dermatology 230:161–169. https://doi.org/10.1159/000369790

83. McLaughlin S, Tobin RJ, Leonard S, et al (2006) The role of digital photography and electronic referral in the triage of patients with suspected skin cancer. Br J Dermatol 154:188–190. https://doi.org/10.1111/j.1365-2133.2005.07009.x

84. Gaudy-Marqueste C, Dubois M, Richard M-A, et al (2011) Cognitive training with photographs as a new concept in an education campaign for self-detection of melanoma: a pilot study in the community. J Eur Acad Dermatol Venereol JEADV 25:1099–1103. https://doi.org/10.1111/j.1468-3083.2010.03940.x

85. Quéreux G, Nguyen J-M, Volteau C, et al (2010) Creation and test of a questionnaire for self-assessment of melanoma risk factors. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP 19:48–54. https://doi.org/10.1097/CEJ.0b013e328333d113

86. van der Geer S, Frunt M, Romero HL, et al (2012) One-stop-shop treatment for basal cell carcinoma, part of a new disease management strategy. J Eur Acad Dermatol Venereol JEADV 26:1154–1157. https://doi.org/10.1111/j.1468-3083.2011.04184.x

87. Moffatt CRM, Green AC, Whiteman DC (2006) Diagnostic accuracy in skin cancer clinics: the Australian experience. Int J Dermatol 45:656–660. https://doi.org/10.1111/j.1365-4632.2006.02772.x

88. Raasch BA (1999) Suspicious skin lesions and their management. Aust Fam Physician 28:466–471

89. Tugrul B, Yalici-Armagan B, Demirdag HG, Gunduz O (2022) Evaluation of Diagnostic Accuracy and Therapeutic Approach of Dermatologists and Plastic Surgeons To Non-Melanocytic Skin Lesions By Using Telemedicine. Dermatol Pract Concept 12:e2022124. https://doi.org/10.5826/dpc.1203a124

90. Whited JD, Hall RP, Simel DL, Horner RD (1997) Primary Care Clinicians' Performance for Detecting Actinic Keratoses and Skin Cancer. Arch Intern Med 157:985–990. https://doi.org/10.1001/archinte.1997.00440300089007

91. Ankad BS, Smitha SV, Koti VR (2020) Basic Science of Dermoscopy. Clin Dermatol Rev 4:69. https://doi.org/10.4103/CDR.CDR_71_20

92. Pan Y, Gareau DS, Scope A, et al (2008) Polarized and nonpolarized dermoscopy: the explanation for the observed differences. Arch Dermatol 144:828–829. https://doi.org/10.1001/archderm.144.6.828

93. Sinz C, Tschandl P, Rosendahl C, et al (2017) Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. J Am Acad Dermatol 77:1100–1109. https://doi.org/10.1016/j.jaad.2017.07.022

94. Nischal U, Nischal KC, Khopkar U (2008) Techniques of Skin Biopsy and Practical Considerations. J Cutan Aesthetic Surg 1:107. https://doi.org/10.4103/0974-2077.44174

95. Sharma KS, Lim P, Brotherston MT (2016) Excision versus incision biopsy in the management of malignant melanoma. J Dermatol Treat 27:88–90. https://doi.org/10.3109/09546634.2015.1034083

96. (2023) Skin Biopsy Specialists | Auckland, NZ | Skintel. https://skintel.co.nz/treatments/skin-biopsy/. Accessed 9 Sep 2024

97. Themstrup L, Banzhaf CA, Mogensen M, Jemec GBE (2014) Optical coherence tomography imaging of non-melanoma skin cancer undergoing photodynamic therapy reveals subclinical residual lesions. Photodiagnosis Photodyn Ther 11:7–12. https://doi.org/10.1016/j.pdpdt.2013.11.003

98. Longo C, Lallas A, Kyrgidis A, et al (2014) Classifying distinct basal cell carcinoma subtype by means of dermatoscopy and reflectance confocal microscopy. J Am Acad Dermatol 71:716-724.e1. https://doi.org/10.1016/j.jaad.2014.04.067

99. Barun VV, Ivanov AP (2009) Estimation of the spectral absorption of light by components of human skin. Opt Spectrosc 106:84–91. https://doi.org/10.1134/S0030400X09010111

100. Bashkatov AN, Genina EA, Tuchin VV (2011) Optical properties of skin, subcutaneous, and muscle tissues: a review. J Innov Opt Health Sci 04:9–38. https://doi.org/10.1142/S1793545811001319

101. Van Gemert MJC, Jacques SL, Sterenborg HJCM, Star WM (1989) Skin optics. IEEE Trans Biomed Eng 36:1146–1154. https://doi.org/10.1109/10.42108

102. Lister T, Wright PA, Chappell PH (2012) Optical properties of human skin. J Biomed Opt 17:090901. https://doi.org/10.1117/1.JBO.17.9.090901

103. Tseng S-H, Bargo P, Durkin A, Kollias N (2009) Chromophore concentrations, absorption and scattering properties of human skin in-vivo. Opt Express 17:14599–14617. https://doi.org/10.1364/OE.17.014599

104. Salomatina EV, Jiang B, Novak J, Yaroslavsky AN (2006) Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range. J Biomed Opt 11:064026. https://doi.org/10.1117/1.2398928

105. Rajaram N, Kovacic D, Migden MF, et al (2009) In vivo determination of optical properties and fluorophore characteristics of non-melanoma skin cancer. In: Photonic Therapeutics and Diagnostics V. SPIE, p 716102

106. Zellweger M (2000) Fluorescence spectroscopy of exogenous, exogenously-induced and endogenous fluorophores for the photodetection and photodynamic therapy of cancer /

107. Borisova EG, Angelova LP, Pavlova EP (2014) Endogenous and Exogenous Fluorescence Skin Cancer Diagnostics for Clinical Applications. IEEE J Sel Top Quantum Electron 20:211–222. https://doi.org/10.1109/JSTQE.2013.2280503

108. Dremin VV, Dunaev AV (2016) How the melanin concentration in the skin affects the fluorescence-spectroscopy signal formation. J Opt Technol 83:43–48. https://doi.org/10.1364/JOT.83.000043

109. Drakaki E, Vergou T, Dessinioti C, et al (2012) Spectroscopic methods for the photodiagnosis of nonmelanoma skin cancer. J Biomed Opt 18:061221. https://doi.org/10.1117/1.JBO.18.6.061221

110. Tseng S-H, Tzeng S-Y, Liaw Y-K, et al (2012) Noninvasive evaluation of collagen and hemoglobin contents and scattering property of in vivo keloid scars and normal skin using diffuse reflectance spectroscopy: pilot study. J Biomed Opt 17:077005. https://doi.org/10.1117/1.JBO.17.7.077005

111. Zonios G, Dimou A, Bassukas I, et al (2008) Melanin absorption spectroscopy: new method for noninvasive skin investigation and melanoma detection. J Biomed Opt 13:014017. https://doi.org/10.1117/1.2844710

112. Nishidate I, Aizu Y, Mishina H (2004) Estimation of melanin and hemoglobin in skin tissue using multiple regression analysis aided by Monte Carlo simulation. J Biomed Opt 9:700–710. https://doi.org/10.1117/1.1756918

113. Mehrübeoğlu M, Kehtarnavaz N, Marquez G, et al (2002) Skin lesion classification using oblique-incidence diffuse reflectance spectroscopic imaging. Appl Opt 41:182–192. https://doi.org/10.1364/AO.41.000182

114. Talari ACS, Movasaghi Z, Rehman S, Rehman I ur (2015) Raman Spectroscopy of Biological Tissues. Appl Spectrosc Rev 50:46–111. https://doi.org/10.1080/05704928.2014.923902

115. Wartewig S (2006) IR and Raman Spectroscopy: Fundamental Processing. John Wiley & Sons

116. Meng H, Yin Y, Wu W, et al (2021) Raman spectroscopic analysis of skin penetration and moisturizing effects of Bionics vernix caseosa cream compared with Vaseline. Technol Health Care 29:327–334. https://doi.org/10.3233/THC-218030

117. Lui H, Zhao J, McLean D, Zeng H (2012) Real-time Raman Spectroscopy for In Vivo Skin Cancer Diagnosis. Cancer Res 72:2491–2500. https://doi.org/10.1158/0008-5472.CAN-11-4061

118. de Oliveira Nunes L, Martin AA, Silveira Jr. L, Zampieri M (2003) FT-Raman spectroscopy study for skin cancer diagnosis. Spectroscopy 17:597–602

119. Lieber CA, Majumder SK, Ellis DL, et al (2008) In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy. Lasers Surg Med 40:461–467. https://doi.org/10.1002/lsm.20653

120. Silveira FL, Pacheco MTT, Bodanese B, et al (2015) Discrimination of non-melanoma skin lesions from non-tumor human skin tissues in vivo using Raman spectroscopy and multivariate statistics. Lasers Surg Med 47:6–16. https://doi.org/10.1002/lsm.22318

121. Canpolat M, Akman-Karakas A, Ayse Gökhan-Ocak G, et al (2012) Diagnosis and Demarcation of Skin Malignancy Using Elastic Light Single-Scattering Spectroscopy: A Pilot Study. Dermatol Surg 38:215. https://doi.org/10.1111/j.1524-4725.2011.02174.x

122. Zakharov VP, Timchenko PE, Timchenko EV, et al (2015) Backscattering spectroscopy for assessing skin tumor. J Biomed Photonics Eng 1:164–168

123. Gareau DS, Karen JK, Dusza SW, et al (2009) Sensitivity and specificity for detecting basal cell carcinomas in Mohs excisions with confocal fluorescence mosaicing microscopy. J Biomed Opt 14:034012. https://doi.org/10.1117/1.3130331

124. Patalay R, Talbot C, Alexandrov Y, et al (2011) Quantification of cellular autofluorescence of human skin using multiphoton tomography and fluorescence lifetime imaging in two spectral detection channels. Biomed Opt Express 2:3295–3308. https://doi.org/10.1364/BOE.2.003295

125. Balu M, Zachary CB, Harris RM, et al (2015) In Vivo Multiphoton Microscopy of Basal Cell Carcinoma. JAMA Dermatol 151:1068–1074. https://doi.org/10.1001/jamadermatol.2015.0453

126. Mogensen M, Nürnberg BM, Thrane L, et al (2011) How histological features of basal cell carcinomas influence image quality in optical coherence tomography. J Biophotonics 4:544–551. https://doi.org/10.1002/jbio.201100006

127. Kuzmina I, Diebele I, Jakovels D, et al (2011) Towards noncontact skin melanoma selection by multispectral imaging analysis. J Biomed Opt 16:060502. https://doi.org/10.1117/1.3584846

128. Ahlgrimm-Siess V, Laimer M, Rabinovitz HS, et al (2018) Confocal Microscopy in Skin Cancer. Curr Dermatol Rep 7:105–118. https://doi.org/10.1007/s13671-018-0218-9

129. (2008) Confocal Microscopy / Denis Semwogerere, Eric R. Weeks. In: Encyclopedia of Biomaterials and Biomedical Engineering, 2nd ed. CRC Press

130. Pellacani G, Cesinaro AM, Seidenari S (2005) *In vivo* assessment of melanocytic nests in nevi and melanomas by reflectance confocal microscopy. Mod Pathol 18:469–474. https://doi.org/10.1038/modpathol.3800330

131. Segura S, Puig S, Carrera C, et al (2009) Development of a two-step method for the diagnosis of melanoma by reflectance confocal microscopy. J Am Acad Dermatol 61:216–229. https://doi.org/10.1016/j.jaad.2009.02.014

132. Guitera P, Pellacani G, Longo C, et al (2009) *In Vivo* Reflectance Confocal Microscopy Enhances Secondary Evaluation of Melanocytic Lesions. J Invest Dermatol 129:131–138. https://doi.org/10.1038/jid.2008.193

133. Waddell A, Star P, Guitera P (2018) Advances in the use of reflectance confocal microscopy in melanoma. Melanoma Manag 5:MMT04. https://doi.org/10.2217/mmt-2018-0001

134. Horn M, Gerger A, Ahlgrimm-Siess V, et al (2008) Discrimination of Actinic Keratoses from Normal Skin with Reflectance Mode Confocal Microscopy. Dermatol Surg 34:620

135. Liu L, Yang Q, Zhang M, et al (2019) Fluorescence lifetime imaging microscopy and its applications in skin cancer diagnosis. J Innov Opt Health Sci 12:1930004. https://doi.org/10.1142/S1793545819300040

136. Patalay R, Talbot C, Munro I, et al (2011) Fluorescence lifetime imaging of skin cancer. In: Photonic Therapeutics and Diagnostics VII. SPIE, pp 55–61

137. Luo T, Lu Y, Liu S, et al (2017) Phasor–FLIM as a Screening Tool for the Differential Diagnosis of Actinic Keratosis, Bowen's Disease, and Basal Cell Carcinoma. Anal Chem 89:8104–8111. https://doi.org/10.1021/acs.analchem.7b01681

138. Miller JP, Habimana-Griffin L, Edwards TS, Achilefu S (2017) Multimodal fluorescence molecular imaging for in vivo characterization of skin cancer using endogenous and exogenous fluorophores. J Biomed Opt 22:066007. https://doi.org/10.1117/1.JBO.22.6.066007

139. Pires L, Nogueira MS, Pratavieira S, et al (2014) Time-resolved fluorescence lifetime for cutaneous melanoma detection. Biomed Opt Express 5:3080–3089. https://doi.org/10.1364/BOE.5.003080

140. Tsai T-H, Jee S-H, Dong C-Y, Lin S-J (2009) Multiphoton microscopy in dermatological imaging. J Dermatol Sci 56:1–8. https://doi.org/10.1016/j.jdermsci.2009.06.008

141. Campagnola PJ, Loew LM (2003) Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. Nat Biotechnol 21:1356–1360. https://doi.org/10.1038/nbt894

142. Zipfel WR, Williams RM, Christie R, et al (2003) Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation. Proc Natl Acad Sci U S A 100:7075–7080. https://doi.org/10.1073/pnas.0832308100

143. Paoli J, Smedh M, Wennberg A-M, Ericson MB (2008) Multiphoton Laser Scanning Microscopy on Non-Melanoma Skin Cancer: Morphologic Features for Future Non-Invasive Diagnostics. J Invest Dermatol 128:1248–1255. https://doi.org/10.1038/sj.jid.5701139

144. Paoli J, Smedh M, Ericson MB (2009) Multiphoton laser scanning microscopy--a novel diagnostic method for superficial skin cancers. Semin Cutan Med Surg 28:190–195. https://doi.org/10.1016/j.sder.2009.06.007

145. Sun TY, Haberman AM, Greco V (2017) Preclinical Advances with Multiphoton Microscopy in Live Imaging of Skin Cancers. J Invest Dermatol 137:282–287. https://doi.org/10.1016/j.jid.2016.08.033

146. Jørgensen TM, Tycho A, Mogensen M, et al (2008) Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography. Skin Res Technol 14:364–369. https://doi.org/10.1111/j.1600-0846.2008.00304.x

147. Forsea A, Carstea E, Ghervase L, et al (2010) Clinical application of optical coherence tomography for the imaging of non–melanocytic cutaneous tumors: a pilot multi–modal study. J Med Life 3:381–389

148. Boone MALM, Norrenberg S, Jemec GBE, Del Marmol V (2013) Imaging actinic keratosis by high-definition optical coherence tomography. Histomorphologic correlation: a pilot study. Exp Dermatol 22:93–97. https://doi.org/10.1111/exd.12074

149. Hinz T, Ehler L-K, Voth H, et al (2011) Assessment of Tumor Thickness in Melanocytic Skin Lesions: Comparison of Optical Coherence Tomography, 20-MHz Ultrasound and Histopathology. Dermatology 223:161–168. https://doi.org/10.1159/000332845

150. Gambichler T, Regeniter P, Bechara FG, et al (2007) Characterization of benign and malignant melanocytic skin lesions using optical coherence tomography in vivo. J Am Acad Dermatol 57:629–637. https://doi.org/10.1016/j.jaad.2007.05.029

151. Korde VR, Bonnema GT, Xu W, et al (2007) Using optical coherence tomography to evaluate skin sun damage and precancer. Lasers Surg Med 39:687–695. https://doi.org/10.1002/lsm.20573

152. Mogensen M, Joergensen TM, Nürnberg BM, et al (2009) Assessment of Optical Coherence Tomography Imaging in the Diagnosis of Non-Melanoma Skin Cancer and Benign Lesions Versus Normal Skin: Observer-Blinded Evaluation by Dermatologists and Pathologists. Dermatol Surg 35:965. https://doi.org/10.1111/j.1524-4725.2009.01164.x

153. Strasswimmer J, Pierce MC, Park H, et al (2004) Polarization-sensitive optical coherence tomography of invasive basal cell carcinoma. J Biomed Opt 9:292–298. https://doi.org/10.1117/1.1644118

154. Baronti S, Casini A, Lotti F, Porcinai S (1998) Multispectral imaging system for the mapping of pigments in works of art by use of principal-component analysis. Appl Opt 37:1299–1309. https://doi.org/10.1364/AO.37.001299

155. Gevaux L (2019) 3D-hyperspectral imaging and optical analysis of skin for the human face. Phdthesis, Université de Lyon

156. Mathews SA (2008) Design and fabrication of a low-cost, multispectral imaging system. Appl Opt 47:F71–F76. https://doi.org/10.1364/AO.47.000F71

157. Tomatis S, Bono A, Bartoli C, et al (2003) Automated melanoma detection: Multispectral imaging and neural network approach for classification. Med Phys 30:212–221. https://doi.org/10.1118/1.1538230

158. Carrara M, Bono A, Bartoli C, et al (2007) Multispectral imaging and artificial neural network: mimicking the management decision of the clinician facing pigmented skin lesions. Phys Med Biol 52:2599. https://doi.org/10.1088/0031-9155/52/9/018

159. Diebele I, Kuzmina I, Lihachev A, et al (2012) Clinical evaluation of melanomas and common nevi by spectral imaging. Biomed Opt Express 3:467–472. https://doi.org/10.1364/BOE.3.000467

160. Neittaanmäki-Perttu N, Grönroos M, Tani T, et al (2013) Detecting field cancerization using a hyperspectral imaging system. Lasers Surg Med 45:410–417. https://doi.org/10.1002/lsm.22160

161. Johansen TH, Møllersen K, Ortega S, et al (2020) Recent advances in hyperspectral imaging for melanoma detection. WIREs Comput Stat 12:e1465. https://doi.org/10.1002/wics.1465

162. Leon R, Martinez-Vega B, Fabelo H, et al (2020) Non-Invasive Skin Cancer Diagnosis Using Hyperspectral Imaging for In-Situ Clinical Support. J Clin Med 9:1662. https://doi.org/10.3390/jcm9061662

163. Courtenay LA, González-Aguilera D, Lagüela S, et al (2021) Hyperspectral imaging and robust statistics in non-melanoma skin cancer analysis. Biomed Opt Express 12:5107–5127. https://doi.org/10.1364/BOE.428143

164. Zherdeva LA, Bratchenko IA, Myakinin OO, et al (2016) In vivo hyperspectral imaging and differentiation of skin cancer. In: Optics in Health Care and Biomedical Optics VII. SPIE, pp 658–665

165. Al-Arashi MY, Salomatina E, Yaroslavsky AN (2007) Multimodal confocal microscopy for diagnosing nonmelanoma skin cancers. Lasers Surg Med 39:696–705. https://doi.org/10.1002/lsm.20578

166. Salomatina-Motts E, Neel VA, Yaroslavskaya AN (2009) Multimodal polarization system for imaging skin cancer. Opt Spectrosc 107:884–890. https://doi.org/10.1134/S0030400X0912008X

167. Heuke S, Vogler N, Meyer T, et al (2013) Detection and Discrimination of Non-Melanoma Skin Cancer by Multimodal Imaging. Healthcare 1:64–83. https://doi.org/10.3390/healthcare1010064

168. Díaz S, Krohmer T, Moreira Á, et al (2019) An Instrument for Accurate and Non-Invasive Screening of Skin Cancer Based on Multimodal Imaging. IEEE Access 7:176646–176657. https://doi.org/10.1109/ACCESS.2019.2956898

169. Fan B, Neel VA, Yaroslavsky AN (2017) Multimodal imaging for nonmelanoma skin cancer margin delineation. Lasers Surg Med 49:319–326. https://doi.org/10.1002/lsm.22552

170. Borisova E, Troyanova P, Pavlova P, Avramov L (2008) Diagnostics of pigmented skin tumors based on laser-induced autofluorescence and diffuse reflectance spectroscopy. Quantum Electron 38:597. https://doi.org/10.1070/QE2008v038n06ABEH013891

171. Borisova E, Pavlova P, Troyanova, Avramov L (2012) Optical biopsy of human skin - A tool for cutaneous tumours' diagnosis. Int J Bioautomotion Vol 16 Issue 1 2012 Pages 53-72

172. Rajaram N, Reichenberg JS, Migden MR, et al (2010) Pilot clinical study for quantitative spectral diagnosis of non-melanoma skin cancer. Lasers Surg Med 42:876–887. https://doi.org/10.1002/lsm.21009

173. Zakharov VP, Bratchenko IA, Myakinin OO, et al (2014) Combined Raman spectroscopy and autofluoresence imaging method for in vivo skin tumor diagnosis. In: Ultrafast Nonlinear Imaging and Spectroscopy II. SPIE, pp 19–26

174. Bratchenko IA, Khristoforova YA, Bratchenko LA, et al (2021) Optical Biopsy of Amelanotic Melanoma with Raman and Autofluorescence Spectra Stimulated by 785 nm Laser Excitation. J Biomed Photonics Eng 7:020308. https://doi.org/10.18287/JBPE21.07.020308

175. Moy AJ, Feng X, Nguyen HTM, et al (2017) Spectral biopsy for skin cancer diagnosis: initial clinical results. In: Photonics in Dermatology and Plastic Surgery. SPIE, pp 21–26

176. Tuzlukov V (2018) Signal Processing Noise. CRC Press

177. Barton SJ, Ward TE, Hennelly BM (2018) Algorithm for optimal denoising of Raman spectra. Anal Methods 10:3759–3769. https://doi.org/10.1039/C8AY01089G

178. Singh K, Upadhyaya DS (2012) Outlier Detection: Applications And Techniques. 9:

179. Rinnan Å, Berg F van den, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

180. Komer B, Bergstra J, Eliasmith C (2014) Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-Learn. Austin, Texas, pp 32–37

181. Rodriguez-Galiano VF, Luque-Espinar JA, Chica-Olmo M, Mendes MP (2018) Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. Sci Total Environ 624:661–672. https://doi.org/10.1016/j.scitotenv.2017.12.152

182. Chandrashekar G, Sahin F (2014) A survey on feature selection methods. Comput Electr Eng 40:16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

183. Jović A, Brkić K, Bogunović N (2015) A review of feature selection methods with applications. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp 1200–1205

184. Anowar F, Sadaoui S, Selim B (2021) Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput Sci Rev 40:100378. https://doi.org/10.1016/j.cosrev.2021.100378

185. Sorzano COS, Vargas J, Montano AP (2014) A survey of dimensionality reduction techniques

186. Huang X, Wu L, Ye Y (2019) A Review on Dimensionality Reduction Techniques. Int J Pattern Recognit Artif Intell 33:1950017. https://doi.org/10.1142/S0218001419500174

187. Jolliffe I (2011) Principal Component Analysis. In: Lovric M (ed) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg, pp 1094–1096

188. Abdi H, Williams LJ (2010) Principal component analysis. WIREs Comput Stat 2:433–459. https://doi.org/10.1002/wics.101

189. Rosipal R, Girolami M, Trejo LJ, Cichocki A (2001) Kernel PCA for Feature Extraction and De-Noising in Nonlinear Regression. Neural Comput Appl 10:231–243. https://doi.org/10.1007/s521-001-8051-z

190. Soman KP, LOGANATHAN R, AJAY V (2009) Machine Learning with SVM and Other Kernel Methods. PHI Learning Pvt. Ltd.

191. Lee J, Verleysen M (2007) Nonlinear Dimensionality Reduction

192. Yang M-H (2002) Extended Isomap for classification. In: 2002 International Conference on Pattern Recognition. pp 615–618 vol.3

193. Hinton GE, Roweis S (2002) Stochastic Neighbor Embedding. In: Advances in Neural Information Processing Systems. MIT Press

194. Maaten L, Hinton GE (2008) Visualizing Data using t-SNE. J Mach Learn Res

195. McInnes L, Healy J, Melville J (2020) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

196. Murphy BW, Webster RJ, Turlach BA, et al (2005) Toward the discrimination of early melanoma from common and dysplastic nevus using fiber optic diffuse reflectance spectroscopy. J Biomed Opt 10:064020. https://doi.org/10.1117/1.2135799

197. Li L, Zhang Q, Ding Y, et al (2014) Automatic diagnosis of melanoma using machine learning methods on a spectroscopic system. BMC Med Imaging 14:36. https://doi.org/10.1186/1471-2342-14-36

198. Lenhardt L, Zeković I, Dramićanin T, Dramićanin MD (2013) Artificial neural networks for processing fluorescence spectroscopy data in skin cancer diagnostics. Phys Scr 2013:014057. https://doi.org/10.1088/0031-8949/2013/T157/014057

199. Dietterich TG (2000) Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–15

200. Sen PC, Hajra M, Ghosh M (2020) Supervised Classification Algorithms in Machine Learning: A Survey and Review. In: Mandal JK, Bhattacharya D (eds) Emerging Technology in Modelling and Graphics. Springer, Singapore, pp 99–111

201. Amari S, Wu S (1999) Improving support vector machine classifiers by modifying kernel functions. Neural Netw 12:783–789. https://doi.org/10.1016/S0893-6080(99)00032-5

202. Cheong S, Oh S-H, Lee S-Y (2004) Support Vector Machines with Binary Tree Architecture for Multi-Class Classification

203. Weston J, Watkins C (1999) Support vector machines for multi-class pattern recognition

204. Crammer K, Singer Y (2001) On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines. J Mach Learn Res 2:265–292

205. Lee Y, Lin Y, Wahba G (2004) Multicategory Support Vector Machines: Theory and Application to the Classification of Microarray Data and Satellite Radiance Data. J Am Stat Assoc 99:67–81. https://doi.org/10.1198/016214504000000098

206. Lauer F, Guermeur Y (2011) MSVMpack: A Multi-Class Support Vector Machine Package. J Mach Learn Res 12:2293–2296

207. Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13:415–425. https://doi.org/10.1109/72.991427

208. Mathur A, Foody GM (2008) Multiclass and Binary SVM Classification: Implications for Training and Classification Users. IEEE Geosci Remote Sens Lett 5:241–245. https://doi.org/10.1109/LGRS.2008.915597

209. Zhao H, Lai Z, Leung H, Zhang X (2020) Linear Discriminant Analysis. In: Zhao H, Lai Z, Leung H, Zhang X (eds) Feature Learning and Understanding: Algorithms and Applications. Springer International Publishing, Cham, pp 71–85

210. Tharwat A (2016) Linear vs. quadratic discriminant analysis classifier: a tutorial. Int J Appl Pattern Recognit 3:145–180. https://doi.org/10.1504/IJAPR.2016.079050

211. Wright RE (1995) Logistic regression. In: Reading and understanding multivariate statistics. American Psychological Association, Washington, DC, US, pp 217–244

212. Tolles J, Meurer WJ (2016) Logistic Regression: Relating Patient Characteristics to Outcomes. JAMA 316:533–534. https://doi.org/10.1001/jama.2016.7653

213. Pal M (2005) Random forest classifier for remote sensing classification. Int J Remote Sens 26:217–222. https://doi.org/10.1080/01431160412331269698

214. Quinlan JR (1996) Learning decision tree classifiers. ACM Comput Surv 28:71–72. https://doi.org/10.1145/234313.234346

215. Ramchoun H, Ghanou Y, Ettaouil M, Janati Idrissi MA (2016) Multilayer Perceptron: Architecture Optimization and Training. https://doi.org/10.9781/ijimai.2016.415

216. Noriega L (2005) Multilayer Perceptron Tutorial

217. Bauer E, Kohavi R (1999) An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. Mach Learn 36:105–139. https://doi.org/10.1023/A:1007515423169

218. Kumari S, Kumar D, Mittal M (2021) An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. Int J Cogn Comput Eng 2:40–46. https://doi.org/10.1016/j.ijcce.2021.01.001

219. Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits Syst Mag 6:21–45. https://doi.org/10.1109/MCAS.2006.1688199

220. Stulp F, Sigaud O (2015) Many regression algorithms, one unified model: A review. Neural Netw 69:60–79. https://doi.org/10.1016/j.neunet.2015.05.005

221. Maulud D, Abdulazeez AM (2020) A Review on Linear Regression Comprehensive in Machine Learning. J Appl Sci Technol Trends 1:140–147. https://doi.org/10.38094/jastt1457

222. Brandt S (1999) Linear and Polynomial Regression. In: Brandt S (ed) Data Analysis: Statistical and Computational Methods for Scientists and Engineers. Springer, New York, NY, pp 413–426

223. Heiberger RM, Neuwirth E (2009) Polynomial Regression. In: Heiberger RM, Neuwirth E (eds) R Through Excel: A Spreadsheet Interface for Statistics, Data Analysis, and Graphics. Springer, New York, NY, pp 269–284

224. Melkumova LE, Shatskikh SYa (2017) Comparing Ridge and LASSO estimators for data analysis. Procedia Eng 201:746–755. https://doi.org/10.1016/j.proeng.2017.09.615

225. Zhang Z, Lai Z, Xu Y, et al (2017) Discriminative Elastic-Net Regularized Linear Regression. IEEE Trans Image Process 26:1466–1481. https://doi.org/10.1109/TIP.2017.2651396

226. AI P (2023) 5 Types of Regression and their properties. In: Medium. https://towardsdatascience.com/5-types-of-regression-and-their-properties-c5e1fa12d55e. Accessed 13 Jun 2024

227. Madhulatha TS (2012) An Overview on Clustering Methods

228. Estivill-Castro V (2002) Why so many clustering algorithms: a position paper. ACM SIGKDD Explor Newsl 4:65–75. https://doi.org/10.1145/568574.568575

229. Vogt W, Nagel D (1992) Cluster Analysis in Diagnosis. Clin Chem 38:182–198. https://doi.org/10.1093/clinchem/38.2.182

230. II ML, Hall D, Llinas J (2017) Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition. CRC Press

231. Castanedo F (2013) A Review of Data Fusion Techniques. Sci World J 2013:704504. https://doi.org/10.1155/2013/704504

232. Botchkarev A (2019) Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. Interdiscip J Inf Knowl Manag 14:045–076. https://doi.org/10.28945/4184

233. Tatachar AV (2021) Comparative assessment of regression models based on model evaluation metrics. Int Res J Eng Technol IRJET 8:2395–0056

234. Botchkarev A (2018) Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio

235. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623. https://doi.org/10.7717/peerj-cs.623

236. Refaeilzadeh P, Tang L, Liu H (2009) Cross-Validation. In: LIU L, ÖZSU MT (eds) Encyclopedia of Database Systems. Springer US, Boston, MA, pp 532–538

237. Celisse A (2014) Optimal cross-validation in density estimation with the L^{2}-loss. Ann Stat 42:1879–1910. https://doi.org/10.1214/14-AOS1240

238. Rodriguez JD, Perez A, Lozano JA (2010) Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Trans Pattern Anal Mach Intell 32:569–575. https://doi.org/10.1109/TPAMI.2009.187

239. Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. Chemom Intell Lab Syst 56:1–11. https://doi.org/10.1016/S0169-7439(00)00122-2

240. Colas V, Daul C, Khairallah G, et al (2021) Influence of human epidermal thickness on penetration depth of detected photons in spatially-resolved diffuse reflectance spectroscopy: a numerical study. In: Optical Interactions with Tissue and Cells XXXII. SPIE, pp 26–37

241. Colas V, Blondel W, Khairallah G, et al (2021) Proposal for a Skin Layer-Wise Decomposition Model of Spatially-Resolved Diffuse Reflectance Spectra Based on

Maximum Depth Photon Distributions: A Numerical Study. Photonics 8:444. https://doi.org/10.3390/photonics8100444

242. Firooz A, Sadr B, Babakoohi S, et al (2012) Variation of Biophysical Parameters of the Skin with Age, Gender, and Body Region. Sci World J 2012:e386936. https://doi.org/10.1100/2012/386936